

Critical Failures in the World Happiness Report’s Model of National Life Satisfaction:

An Improved Specification and its Implications for Democracy

Alexander Loewi*

March 10, 2024

Abstract

The World Happiness Report is currently the best known and most widely accepted source of information on global life satisfaction. Its six-variable model of satisfaction is reproduced in textbooks and has been published in the same form by the United Nations since 2012.

In this paper we model country-level average life satisfaction, as reported by the World Happiness Report (WHR), using 18 years of available data. A total of 5,527 variables are considered as potential predictors in the model, and a Lasso-penalized regression is used to select between them. In contrast to the six-variable specification that is the primary reported model for the WHR, we find far more variables with meaningful contributions, very different contributions for the WHR’s six variables including two out of six that can be removed entirely without significant consequences to fit, a strictly better model fit both in and out of sample, and a dramatically different interpretation of the results.

In particular, we find the WHR is missing significant variables in areas including discrimination, education, political power, and more. We also find that while the reported model suggests GDP explains roughly 40% of model variation, we estimate the true direct contribution of economic indicators to be 2.3%.

Additionally, the conclusions of the more thorough model are shown to be in close alignment with two major theories of the generalized prejudice literature, which have been repeatedly successful at explaining modern political outcomes. The nature of the model-discovered variables strongly suggests that individuals high in Social Dominance Orientation and Right Wing Authoritarianism pose direct and substantial threats to national satisfaction. These results are also discussed, in particular with respect to the implications for how national satisfaction can be improved.

1 Introduction

The World Happiness Report (WHR) is currently the best known and most widely accepted source of information on global life satisfaction. Its six-variable model of satisfaction is reproduced in textbooks and has been published by the United Nations since 2012.

The WHR justifies this model by saying “The variables used reflect what has been broadly found in the research literature to explain national-level differences in life evaluations” (Helliwell et. al, 2023). However a thorough search for variables that predict satisfaction among existing international data sets finds a dramatically different set, with similarly different interpretations and policy implications.

We find a much better model fit with fourteen variables in the categories of health and survival, social inclusion, self determination, infrastructure, education, inclusive power, and the monetary economy. These contain several classes of variables, including on discrimination, education, and political power, that the

*Dovecote Institute, Washington DC, 20003, E-mail: aloewi@dovecoteinstitute.org

WHR model omits entirely. These variables also do not contain two out of six of the WHR’s variables, as they were simply not chosen by the variable selection process, and do not improve model fit when included.

In addition, we find that the WHR’s specification appears to overestimate the importance of GDP by approximately 1100%. This alone dramatically changes the character of its conclusions.

In this paper we describe the search process for a better model, explain the resulting model, and consider the paths that are available to increase national satisfaction, based on the model results.

1.1 Background

When attempting to be successful in any endeavor, from personal to global, there is no more fundamental step than defining that success. In the United States, the Declaration of Independence proposes indirectly that success be defined as “life, liberty, and the pursuit of happiness” – but concretely, rigorously measured versions of these criteria are rarely discussed. Posed in all seriousness, a simple question does not return an easy answer: What exactly is the United States attempting to achieve? Or even more grandly, What do we want out of life?

Perhaps the most frequently used quantitative objective is economic productivity, measured by Gross Domestic Product (GDP), despite the fact that this measure was never designed for such a purpose (Sen, Stiglitz, and Fitoussi, 2010). Now multiple generations of economists have pointed out GDP’s fundamental shortcomings as a national guide, and international working groups of leading researchers have been established to find alternatives. Multi-part objectives (for example, the sum of GDP and average lifespan) are an easy solution on one level, and there have been several serious attempts at compiling lists of things that seem like intuitively good goals. However, these lists introduce the equally difficult problem of how to meaningfully weight their many components, which can number into the hundreds. The unavoidable challenge of this approach is that for any such weighting to be meaningful, there must be an objective to weight against.

As of now, the most plausible single-measure objective to use as a national, and even global goal, is subjective life satisfaction (Diener et. al, 2009). Unlike “happiness,” or “positive affect,” which typically refer to a necessarily short-lived emotional state, “satisfaction” in the literature refers to a cognitive judgment about the respondent’s overall satisfaction with the totality of their life. Most commonly, this is measured with Diener’s Satisfaction With Life Survey (SWLS) (Diener et. al, 1985) or the Cantril Ladder (Cantril, 1965), which take a similar approach. A large body of research has demonstrated these measures are stable, meaningful, intuitive, valid, and largely comparable across languages, countries, and cultures (Diener et. al, 2009).

In light of these findings from satisfaction researchers, and combined with the international desire for better accounting of human activity, the consulting firm Gallup Inc. has included questions on self-reported life satisfaction in its World Poll since 2005. This effort uses the Cantril ladder, or life ladder, which asks the following question:

“Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?”

Since 2012, and despite its confusing title, these results on satisfaction have been the central reported substance of the World Happiness Report (Helliwell et. al, 2023) which is released in collaboration with the United Nations.¹

The WHR is now the world’s best known and most widely distributed source of information on global satisfaction. The centerpiece of this report is the WHR’s six-variable model of satisfaction, which uses six

¹While it is currently the case within the field of specialists that “happiness” is used as a shorthand to refer to both “positive affect” (a short term emotion) and “life satisfaction” (a long term cognitive evaluation), whenever there is the potential for confusion, even the Happiness Report itself is usually very careful to specify which one is actually meant, because they are simply different concepts. We find it unfortunate for everyone outside the field however that the title of the report uses a word known almost exclusively as “positive affect,” and the quantitative substance of the report focuses almost exclusively on a concept with a very different meaning. Though we also acknowledge, without malice, that the “World Satisfaction Report” may have far fewer readers.

hand-picked variables in a linear regression to predict national-level satisfaction outcomes. This model has been published in the same form since 2012, is reproduced verbatim in macroeconomics texts (Larrain, 2020), taught in classrooms (Gilman, 2023), and distributed by the United Nations.

However, this flagship model immediately raises concerns. The list of six variables is both surprisingly short, and limited in its contents. Unfortunately, access to the primary data requires an annual subscription of \$30,000 USD (Gallup sales team, personal conversation, 2022), posing a substantial and often definitive barrier for researchers interested in a re-analysis (anonymous economics professor, personal conversation, 2023).

In light of this concern, and the extreme importance of the conclusions, this paper attempts to re-run the model of global life satisfaction by exclusively using publicly, or cheaply available data, and by casting as broad a net as possible for variables that might possibly belong in the model.

The results obtained differ dramatically from those reported by Gallup and the UN.

1.2 The Current WHR Model and Publicly Available Data

While Gallup charges a subscription fee for individual level data, and even for a large collection of summarized data, they do share some summarized data for each of the countries where the survey is conducted. For each participating country, the data that is shared, and used in the reported model, includes:

1. The outcome: average national life satisfaction, as measured by the Cantril Ladder
2. Average log GDP per capita
3. Average healthy life expectancy
4. Average self-reported social support
5. Average generosity, a measure of charitable giving relative to wealth
6. Average level of perceptions of corruption in public and private institutions
7. Average self-reported feelings of freedom

Each of the self-reported variables are reported to have been estimated from around one thousand samples per country (Helliwell et. al, 2023), which is approximately the sample size of standard national polling in the United States.

1.3 Principles for an improved model

1.3.1 The greatest possible scope

The primary motivation of the re-analysis is a skepticism that the published six variables can meaningfully “explain” – the word used repeatedly by the WHR – either individual, or national, life satisfaction. Given the highly correlated nature of many of the independent predictors, the primary concern was not the original model’s ability to predict life satisfaction with low error, but the more serious and structural problem of omitted variables.

The problem of omitted variables is critical because the value of the model, should there be any, is in its ability to allow us to *improve* satisfaction. As such, it must be able to guide policy decisions, at least in part through prioritizing the variables included based on their estimated importance. Consequently, every omitted variable represents the possibility of an ignored problem, as well as a mis-estimation of the included problems. Thus the highest priority was given to specifying the new model as thoroughly and extensively as possible.

In practice, this was carried out through a “kitchen sink” approach. The availability of country-level averages made it possible to find any dataset that was also labeled by country and join it to the provided outcome data. With such an unbounded objective, it is impossible to ever claim a truly comprehensive collection,

but a rigorous search compiled data sets from a large number of reputable sources. These sources included the World Bank, Varieties of Democracy, the Organisation for Economic Co-operation and Development (OECD), the United Nations, the international micro-lending organization Kiva, the Social Progress Imperative, the World Inequality Database, Journalists Without Borders, the International Monetary Fund, the World Health Organization, the Pew Research Center, and many others. In all, the final dataset included variables on extensive social, governmental, technological, environmental, health, financial, and economic indicators, including GDP, economic sector size, individual rights, trade deficits, life expectancy, charitable giving, abortion access, poverty, inequality, corruption, internet access, LGBTQ+ safety, national emotions, freedom of the press, and much more.

The scope and detail of the final collection was considered to be satisfactory, especially in light of the many areas it covers that the WHR’s hand-picked model does not. However, given the extraordinary scope of the original search, there are three areas that were felt to be underrepresented in this compiled data.

The first is arts and culture. The United Nations Educational, Scientific and Cultural Organization (UNESCO) provides national accounting in this area, but currently only for 58 countries. This is far too few to be used in the models of the countries included in the WHR, of which there are typically around 150 per year in the shared data. This is also the only area for which it is considered that no data at all is included.

The second area is racism. This is an inherently difficult variable to collect because of the many different ways that race is understood globally, but systematized national data that attempts to capture it does exist in places such as the World Values Survey. Unfortunately, the current state of that collection also only includes around 60 countries. However, there are multiple variables on non-racial minority discrimination in the compiled data, several of which play prominent roles in the model, which are expected to act as partial proxies in the face of this specific omission.

The third broad area felt to be underrepresented is a detailed breakdown of government spending. Data on this is collected by the International Monetary Fund (IMF) but again, not with sufficient coverage to be included for the full range of activities that governments support. (One prominent exception is spending on education.) However, several hundred variables minutely describing many other aspects of government are available from the Varieties of Democracy dataset.

1.3.2 The Most Objective Possible Specification

In addition to casting a wide net for possible variables, it was considered crucial to follow a maximally data-driven approach to model selection. For a question as large as the one the model poses – What are the drivers of large-scale human satisfaction? – a primarily manual specification was considered unnecessary, badly prone to bias, and far too consequential for the potentially enormous errors of omission. These two principles of scope and objectivity of specification drove the model selection process.

2 Methodology

2.1 The Data

The outcome data was made available by Gallup, as supplemental files to the WHR (Helliwell et. al, 2023). All national average estimates are made available beginning from 2005, for a total of 2,199 country-year measurements over the six WHR variables, of which 1,964 are complete. These complete observations are used by the WHR in its reported model.

In our effort to determine which independent variables should be included in a model, first, over five thousand variables were collected for each of the years 2019 and 2023.

The year 2019 was chosen because it was recent, but also before COVID, after which patterns might have become exceptional. The year 2023 was chosen because it could accommodate the most recently available data, including several key datasets that did not have historical archives. In the cases where a source did not

have up to date 2023 data, the most recent available year was used when necessary. After multiple means of examination, these two years were considered to be satisfactorily representative of the others.

The results from these initial models suggested that of the multiple dozen identified, only five datasets contained variables that were selected by the algorithm (or, contained a superset of the variables available from other sources). These datasets were the WHR (WHR/Gallup, 2023), the Sustainable Development Goals (SDG, 2023), the Social Progress Imperative (SPI, 2023), Varieties of Democracy (V-Dem) (Coppedge et. al, 2023), and World Bank data on unemployment (World Bank, 2023).

For the majority of variables, these five datasets included measurements for the full historical range of the WHR data. However there were a number of critical exceptions. Believed to be most important is a variable from the Social Progress Imperative, “Acceptance of gays and lesbians (proportion of pop.)”.² This is the sole public variable found on social acceptance of any LGBTQ+ people over a sufficiently large number of countries and years to be modeled, and it is only available from 2011 onwards.

When restricting analysis to just these years, this variable on social acceptance was so reliably selected by the model that it was chosen even in a specification that was forced to produce as few as six variables, meaning it was selected even before “Freedom to make life choices,” and estimated in this minimal specification at almost double the contribution of GNI per capita. This six-variable data-selected model also matches the WHR’s six-variable model almost exactly for model fit, though it will be shown this is far from exceptional.

Due to our fundamental focus on avoiding the risks of omitted variables, the choice was thus made to model only years from 2011 to the present so that social acceptance of gays and lesbians could be included. While this does exclude several years of WHR data, the relatively small number of countries surveyed during these early years means that we are still using a similar number of observations as the WHR in the model that we present as our central result (1,633 observations instead of 1,964, or 83%). Our selection constitutes twelve years of contiguous, extensively-featured, and contemporary, observations, extending back to before the publication of the first World Happiness Report.

For exhaustive rigor, we also construct models built on the full historical range of available data, though with the strong caveat that it is known not to include core predictors. Still, this model gave very similar results to the model we report, with similarly high accuracy, and choosing very similar variables. Small variations included a greater emphasis on specific diseases – cancer, and anemia in women – and a number of small political variables including regime duration and opposition group size. Based on this substantial similarity, we consider our data to be satisfactorily representative.³

2.2 The Model Estimation Process

The primary challenge for modelling was the large number of independent variables compared to the far smaller number of observations. To solve this problem, the highly under-determined system was first estimated using an iterative procedure centered around a Lasso-penalized linear regression (Tibshirani, 1996), and the heuristic (ultimately relevant only for the longitudinal models) that there should be no fewer than ten observations per independent variable in the final model. The regularization penalty was thus tuned to ensure this. However, because the Lasso produces biased estimates (Wüthrich and Zhu, 2023), after the variables had been selected in this manner, they were used to specify an un-penalized linear model. These results are reported.

2.3 Robustness Checks in the Estimation Process

Given concerns about stability in this highly under-determined system, there were extensive checks for robustness. The main procedure estimated each penalized model 25 times, each time using different random

²While we would ideally have a variable that measured population acceptance of all LGBTQIA individuals, the data available asks specifically about gay and lesbian acceptance. We label and discuss it accordingly for full accuracy and transparency.

³Another important difference was that several of the variables with low significance in our reported model displayed much higher significance when using all 2,199 available observations.

subsets of 95% of the rows. As expected, each such iteration selected somewhat different variables. However, when this process was repeated many times, the results stabilized into a subset of variables that appeared with high frequency. These high-frequency variables were then chosen to include in the next phase model.

In tests, this procedure was found to be far more efficient than a simple application of the Lasso. While the reported model has 14 independent variables, the simple Lasso required 23 variables to achieve a model with the same holdout error. This appeared to be due in part to the Lasso’s propensity to include multiple variables of a similar nature, e.g. “life expectancy at birth” and “life expectancy at 60.” The iterative procedure did a better job of choosing just one. Though in both cases, the frequent appearance of conceptual clusters was considered a strong indication that the model was capturing a stable signal in the data.

2.4 Variable Selection

Several models were fit. The primary model, which we will examine in the greatest detail, is a pooled model with fixed effects for years. This is the same form as the WHR’s primary six-variable model. The goal of our model was a parsimonious and interpretable specification that could meaningfully improve upon the WHR model, but still be in a form where they could be compared.

Other models include a longitudinal model where the goal was to assess the historical consistency of this pooled model, a pure prediction model, where the goal was to determine a theoretical minimum for model error, and a train/validation/test model to ensure that the search procedure was not overfitting.

For the main model, the weights on the penalty took the form $\exp(-\lambda)$. A value of $\lambda = 2.5$ was found to effectively balance prediction power and parsimony. This model produced 16 variables that were selected in 100% of the 25 robustness bootstraps described above. Two of these were removed, without affecting fit, because they were conceptually identical to other variables in the model, and one was replaced with a different conceptually similar variable that showed a better overall fit. These 14 variables are reported.

The pooled specification was used to assess the longitudinal model. This revealed that with just three years of training data, predictions on the remaining nine years of data perfectly matched the out-of-sample accuracy of the full pooled model. This suggests a high stability across time as well.

There was a fear that the Lasso procedure might be testing too many models, and thus run a risk of overfitting. To evaluate this concern, 20% of the data was held out from the search procedure, and used only to assess the model’s true out-of-sample accuracy. This was found to closely match the model’s internal cross-validation accuracy, suggesting the concerns were unfounded. This also seems entirely plausible given the use of only 14 variables to model almost two thousand observations with a simple linear model. The number of parameters available structurally makes the possibility of over-fitting relatively low.

Last, a model was fit with elasticnet (a combination of L1 and L2 penalties) to find a theoretical minimum error. Using the optimized values of $\lambda = 2.9$ and $\alpha = 0.89$, where α is the proportion of the lasso in the overall penalty, a specification was found that had 42 variables, and a cross-validation error of 0.45. Especially given that the outcome scale is 0 to 10, we consider that this minimum is also within an almost negligible distance of the error of the reported model, which is 0.498, but attained using 28 fewer variables that displayed a very similar pattern of categories. This is the reason we do not use more variables.

The reason that we do not use *fewer* than 14 variables is that, in addition to the predicted reduction in model fit, using only a very small number of variables fails to capture important signals that appear in just slightly larger models. Variables on education, governance, and other categories, did not appear in the most strict models, but significantly improved model fit.

Together, these findings give us confidence that the reported specification strikes a strong balance between accuracy, parsimony, and explanatory power. They also make it clear that there is a very substantial space of predictors beyond six with calculable contributions to prediction. Efforts that are not trying to prioritize parsimony so heavily could easily draw on these extended models.

2.5 Handling Missing Data

Variables with missing values were an issue in all data sources, including those from the WHR itself.

Once examined closely, it became clear that less than a quarter of collected variables from 2019 and 2023 were over 75% complete for the 156 countries included in the WHR. It was also observed that allowing variables up to 75% missing would dramatically increase the number of variables available – as, with so many independent variables, practically no observations were entirely complete. Consequently, using 25% missing as a threshold, all of these approximately 1,300 variables were used in the initial models, with the missing values imputed by random forest. After repeated tests, the randomness of the imputation did not appear to play any role in the final results.

Once the key data sets and time periods had been identified, there were 1,058 variables that were available for all 12 years of data, and that were also at least 75% non-missing during those periods. The variable selection process chose between these 1,058.

While 25% missing was allowed in the variables that were searched, in the final pooled specification, the average missingness per column was less than 7%. Tests were performed to assess the impact of the missing data, including removing rows with high missingness, and entirely removing the most missing variables entirely. Neither change made any meaningful difference to the model estimates for the remaining variables.

One exceptional case is data on global poverty and inequality. This was *not* used in any of the reported models, because the availability of this data falls far below the completeness threshold used here – a point also noted explicitly by the WHR as their reason for not incorporating data on inequality. However, we considered these sufficiently critical variables that we attempted to incorporate them in an exploratory analysis. This was done by taking the average of all available measurements for a country in the period under study, and using that “average” (often just a single value) for all modeled years of data. Using this approach, the median household income in a country was chosen in 100% of the 25 model-search bootstraps. This underscores the importance of the empirical fact that the actual income of a typical household can diverge substantially from GDP per capita Nolan, Roser, and Thewissen (2016). The remainder of the selected variables in this specification were highly similar to the model we do report, as was overall model accuracy. We note that measures of inequality, such as a Gini index, were also available but not selected.⁴

—>

3 Results

3.1 The Model

The pooled model, using twelve years of data, is the main result of this paper. This is the model we compare with the current WHR six-variable model, and present as an alternative. We will refer to this as the Dovecote model. The full version contains fourteen variables, in addition to an intercept and fixed effects for each year, which are also used in the WHR models.

Each of these fourteen variables is reported because they were selected by the Lasso and bootstrap process described above in 100% of the iterations, with two exceptions.

Social and political power for women is used in place of the modeled proportion of women employed in the “service” sector, though this sector is defined broadly and includes white-collar jobs. This exchange resulted in an identical model fit, and a very similar measured contribution to model variation, but the replacement

⁴As will be seen more clearly in a close examination of the reported model, the primary emphasis appears to be on first-order, fundamental needs. While a measure on inequality can capture a difference in levels of these basic resources, the model suggests that the more important problem is, some people simply don’t have these resources at all – and whether or not someone else does, is secondary. Though we caution this is both based on limited data, and is only an interpretation.

has a strong increase in significance.⁵ We also interpret the natures of the two variables – social/political power, in place of labor/employment power – to be reasonably similar.⁶

Similarly, years of available pre-primary education is used in place of the years of compulsory education. In this case as well, the measured contributions are similar, and the meaning is closely related, but the replacement is much more strongly significant.⁷

In addition to these two exchanges, two redundant variables were removed outright. GDP per capita and GNI per capita were both selected, but in the interest of parsimony, each was tested, and the fit with GNI was marginally better. When using only one of these variables, it clearly absorbed the contribution of the other, so we believe this makes no meaningful difference. Similarly, life expectancy at birth and life expectancy at sixty were both chosen by the model, but we retain only life expectancy at birth for simplicity and ease of interpretation. No other changes were made, to stay as close as possible to a purely data-driven specification.

With several very large contributions, and several contributions more than an order of magnitude smaller, this model also strongly suggests the rapidly diminishing marginal effect of additional variables. While many other variables were manually added to the model in exploratory analyses, none ever approached contributions on the scale of the largest variables. This pattern suggests that the most important effects, at least among the available 1,058 variables, have been found.

Table 1 shows the Dovecote model. Following the WHR, robust standard errors are used, clustered by country.

Also following the WHR’s format for reporting, Figure 1 displays the average contribution of each variable in the Dovecote model to the predictions. Bars are colored by significance level, where navy indicates significance at the .001 level, dark blue at the .01 level, sky blue at the .05 level, and light blue at the .10 level. As with other technical details, we follow the WHR’s convention of significance levels. When reading this table however it is important to remember that the variables were selected not primarily for their significance, but for their ability to predict satisfaction in an out-of-sample model. (And that this significance is reliably much higher when using data going back to 2005.)

3.2 Format and interpretation of the reported results

While the estimated coefficients and fit statistics are included in the WHR, the report typically gives more space to presenting not the regression coefficients $\hat{\beta}_i$, but rather the “contribution” of each variable in each country, or $\hat{\beta}_i \bar{x}_{ij}$. This is referred to as the portion of a country’s satisfaction “explained by” each variable. For ease of comparison and interpretation we follow the same convention in our plots, but using averages over all countries for simplicity of presentation.

For an interpretation of the magnitudes, consider 1.5 points on the 11-point satisfaction scale to be equal to the nadir of one personal tragedy. This is approximately equal to the average measured drop in life satisfaction for an individual between the beginning of the decline, and the final experience, of the death of a life partner (Infurna et. al, 2020).

In other words, while the individual may (and may not) rebuild their life in the wake of this loss, and increase their satisfaction again later, the average studied individual has dropped approximately 1.5 points at the time of the actual death. Furthermore, it takes an average of five years from this low point to fully recover to the initial level of satisfaction – a measurement on average first taken ten fully years earlier, at the beginning of the partner’s decline.

⁵Though this may be a function more of sample size than actual importance. In a model using all 2,199 observations, both variables were strongly significant, even when both were in the model simultaneously.

⁶While the “service” sector is not always seen as high-powered, the limited number of categories used by the variable means that this variable can also be interpreted as “women employed, but not in agriculture.” We believe this accounts for its ability to effectively proxy for more general social power.

⁷As above, these variables both show strong and simultaneous significance when using all 2,199 observations.

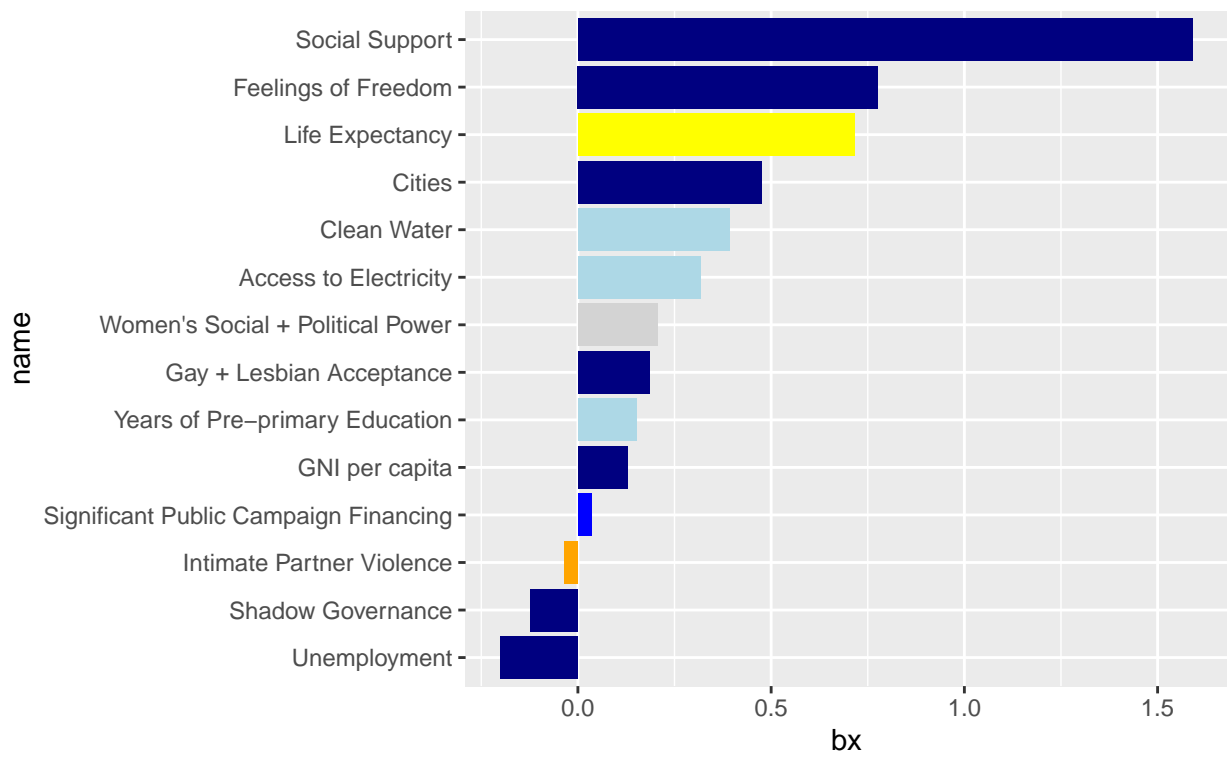


Figure 1: Contributions of the Dovecote Model

	Dovecote Model
GNI per capita	0.00*** (0.00)
Urban Population	0.01*** (0.00)
Social Support	1.97*** (0.29)
Gay + Lesbian Acceptance	0.59*** (0.16)
Clean Water	0.57* (0.31)
Significant Public Campaign Financing	0.06** (0.02)
Shadow Governance	-0.42*** (0.13)
Unemployment	-0.03*** (0.01)
Women's Social + Political Power	0.24 (0.20)
Years of Preprimary Education	0.04* (0.02)
Feelings of Freedom	1.02*** (0.28)
Life Expectancy	0.01 (0.01)
Access to Electricity	0.00* (0.00)
Intimate Partner Violence	-0.00 (0.01)
Adjusted R ²	0.811
Cross Validation Error	0.497
N Obs.	1,633

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 1: The Data-Driven Model, using Year Effects and Robust Standard Errors

Thus while a 0.15 point drop in national satisfaction may look small as a number, it should be interpreted as 1 in 10 people living a life that is worse by the magnitude of a full-blown tragedy. Thus even small numbers can be enormously significant from a behavioral perspective, both in the intensity and duration of effects.⁸

3.3 Variables with multiple variations

Many of the variables in the data were very similar in substance, such as GNI and GDP, or Childhood Undernourishment and Population Undernourishment. When both were selected by the model, the unweighted model would be re-run using each of the alternatives alone to determine which produced the best fit. Here we report the variables that were selected in this manner, and the variables against which they were compared.

3.3.1 GDP and GNI

In different model searches, GNI and GDP per capita were both frequently selected by the Lasso procedure. However, when both are included in the reported specification, they both lose their statistical significance. When testing them one by one, the cross-validation error with GNI is marginally lower. While GNI is better in this respect, it should be no surprise that the two quantities are similar as the numbers themselves are typically very similar for a given country.

There is also the intuitive reason to expect GNI's greater contribution, as it measures the amount of value that comes into a country, not the amount that is produced by it (even if it suffers fundamentally from an inaccurate accounting of that value, in the same ways that GDP does).

A much greater effect on the model comes from the choice to use a logarithmic transformation on these economic indicators, as is done in the WHR's hand-picked model, but not in data-driven Dovecote model. This will be addressed in greater detail below.

3.3.2 Health

While it is one of the most intuitively important variables, estimates of health, lifespan, and disease were surprisingly unpredictable. Over multiple models of different specifications, several diseases were selected by the procedure, including tuberculosis, cancer, HIV, respiratory illness, and anemia, each of which often produced contributions of a comparable magnitude to the variables on lifespan. However because of this broad variety, and their somewhat inconsistent appearance, the decision was made to stick exclusively to variables on lifespan, for reliability, coherence, and parsimony.

Within these variables, the primary model chose both life expectancy at birth and life expectancy at age 60. Life expectancy at birth was manually selected as the single representative of lifespan, simply due to its being a more intuitive measure.

It should be noted that this variable is *not* the same as the WHR's *healthy* life expectancy at birth (HLE), which differs on average from unconditional life expectancy by almost a decade.⁹ HLE was not selected by the procedure, and when it was manually placed in the model it showed almost no contribution. Though even unconditional life expectancy, despite its very large estimated contribution, does not have a statistically significant parameter estimate in this model.

Should this be surprising, we note this is also consistent with the WHR's own pooled model, where HLE does not show as strong significance as its other variables (meaning, significance at the .01 level, instead of the .001 level). This may be due in part to the individual years of 2020 and 2021, in which the parameter for HLE can actually be seen to flip in sign – possibly reflecting the greater risk of COVID-19 for people higher in age.

⁸This finding also begins to hint at the extraordinary range of experiences possible for an individual, as well as the consistent historical difficulty in measuring effects from small interventions.

⁹The preference for an unconditional quantity seems consistent with the pattern of discovered variables, which in general appear to include the most simple and direct rudiments of life. The subjective decision necessary to define “healthy” also has more potential issues than a simple record of mortality.

3.3.3 Shadow governance

One of the more unexpected variables was from the V-Dem dataset, from the column “HOS [head of state] control over.” This was a variable that contained an expert answer to the following question: “In practice, from which of the following bodies must the head of state customarily seek approval prior to making important decisions on domestic policy?” Possible answers included options such as “A foreign power,” “The military,” or “A religious body.” In other words, these variables attempt to measure the degree to which there is functionally a shadow government in a given country, that rules in place of the putative head of state.

As multiple options could be chosen, each selection was encoded as a separate variable, and the Lasso selected from among these. In both annual and pooled models, it frequently selected two specific responses: “None,” which had a positive effect on satisfaction, and “The ruling party or party leadership body (in a one-party system)” which had a strong negative effect on satisfaction.

While both variables appeared frequently, the variable representing the answer “The ruling party or party leadership body (in a one-party system)” was chosen more reliably by the model, and demonstrated stronger significance. Consequently, this is the variable we report in the pooled model. However, we have empirical reason to believe that by itself, it represents a narrower perspective than that reflected by the full data.

3.3.4 Women’s social and political power

In general, composite variables were found to give highly unpredictable results, and were removed. Even when well behaved, the inability to know with certainty which component was responsible for a measured effect was considered to render them unusable. The one notable exception to this was the V-Dem variable “Women political participation index.” This is a composite of only two other variables that are closely related conceptually – the number of women in the country’s legislature, and the amount of power that women hold generally in society. As it had appeared strongly in other models, it was tested as a replacement for the more narrow variable on the proportion of women employed in the service sector, and found not only to improve model fit, but to display a substantial contribution and greater statistical significance. When included in the model with data back to 2005, instead of just 2011, the significance substantially increases again, both for the variable on political power, and for the variable on service sector employment.

3.3.5 Education / information

One major surprise while aggregating data was the inconsistent availability of good enrollment data, despite the literally thousands of columns set aside for education in public data sets. Because these columns were overwhelmingly empty, and updated infrequently, it was necessary to use more easily measured quantities for the pooled models. Three of these available variables were selected with high frequency: the number of years of pre-primary education as described by a country’s laws (not necessarily as practiced), the number of years of compulsory schooling, again as described in law, and the government expenditure on education as a percentage of GDP.

Of these three, the number of years of compulsory education was selected in 100% of the procedure’s iterations, with the number of years of pre-primary education coming in only slightly behind. However pre-primary education showed stronger significance, and was thus included in the model.

3.3.6 Shelter / access to electricity

Early runs of the model selected the “Shelter” composite, created by the Social Progress Imperative. This variable contains components on access to electricity, the number of deaths from household air pollution, the use of clean fuels and technologies for cooking (due to their importance for household air pollution), and dissatisfaction with housing affordability. While these are all important variables with substantial research behind them, given the separate relevance to health, infrastructure, and affordability, the composite was not

considered to be interpretable. It was manually removed, and access to electricity was selected only because the model chose it independently. Electricity was also seen to have the largest contribution when the four components were manually added to the model independently. Still, the model’s choice of the composite gives us reason to believe that the other variables bundled with electricity have important contributions to satisfaction.

In particular, access to electricity may even itself be a partial proxy for the use of clean fuels and technologies for cooking – the importance of which is known to be enormous, and specifically through its mechanism of reducing indoor particulates, and respiratory illnesses. We mention this here because access to electricity did not have sufficient observations to be used in models going all the way back to 2005 – and instead, that model placed a much higher priority on clean cooking, and the negative effects of cooking with coal and biomass.

3.3.7 Variables with no comparisons

Lastly, several variables had no clear analogues. These were social support, feelings of freedom, satisfaction with the quality of water, the proportion of the nation living in cities, the degree of gay and lesbian social acceptance, unemployment, and significant public funding for political campaigns.

3.3.8 Other frequently selected variables

In addition to the variables with very similar meanings, there was another class of variables that showed up frequently in exploratory models, but were not included in the final model. This is either because the pooled model did not select them 100% of the time, or because they showed up only in annual models, or models with larger numbers of variables. However, these variables are still considered of substantial interest for further study.

Included in this set are variables on information access, internet access, press freedom, digital governance, median individual income, lead exposure, the use of restrictions on candidates running for office, civil rights, academic freedom, free and fair elections, online harassment, the major uses of social media, and the use of elected, as opposed to appointed, regional governments – though in this unbounded space of variables, there are doubtless many more worth investigating.

3.4 Comparison with the WHR’s hand-picked model

While every annual publication of the WHR includes multiple specifications of a model for average national satisfaction, these models always retain the six hand-picked variables as a base, and none come close to the specification presented here. It is also these six variables that are consistently included in the major results of the report. The 2023 WHR devoted an entire full-page illustration to each of these six variables separately, simply to emphasize their centrality to the current paradigm. We thus focus our attention on this same set.

There are several immediately visible differences between the Dovecote model estimated here, and the WHR’s flagship hand-picked model. A comparison of the two can be seen in Table 2

3.4.1 Number of variables

The first evident difference is that our model uses fourteen variables (excluding the intercept and fixed effects for year, which are also used by the WHR). While fourteen is still a very small number compared to the presumably thousands of factors that drive the many intricacies of national life satisfaction, it is not at all surprising that a thorough search would reveal a larger number of variables than six. As discussed above, this number is considered to be on optimal tradeoff between parsimony, accuracy, and expressiveness.

	Dovecote	WHR
GNI per capita US	0.00*** (0.00)	
Urban population	0.01*** (0.00)	
Gay + lesbian acceptance	0.59*** (0.16)	
Clean water	0.57* (0.31)	
Significant public campaign financing	0.06** (0.02)	
Shadow wovernment	-0.42*** (0.13)	
Unemployment	-0.03*** (0.01)	
Women's social and political power	0.24 (0.20)	
Years of preprimary education	0.04* (0.02)	
Life expectancy	0.01 (0.01)	
Access to electricity	0.00* (0.00)	
Intimate partner violence	-0.00 (0.01)	
Social support	1.97*** (0.29)	2.53*** (0.36)
Feelings of freedom	1.02*** (0.28)	1.33*** (0.30)
Healthy life expectancy at birth		0.03** (0.01)
Log GDP per capita		0.36*** (0.07)
Perceptions of corruption		-0.72** (0.26)
Generosity		0.54* (0.26)
Adj. R ²	0.810	0.757
CVE	0.498	0.569
Num. obs.	1,633	1,964

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 2: The Dovecote and WHR Models

3.4.2 Model fit

The second major difference with the WHR is model fit. The most recent reported adjusted R-squared for the hand-picked model is 0.76. The adjusted R-squared for the Dovecote model is 0.81.

However, due to concerns of possible over-fitting, out-of-sample cross-validation accuracy was considered more important than in-sample fit, and so the WHR’s six-variable model was reproduced to assess it on these grounds. As confirmation that this estimate is valid, we observed that the coefficient estimates, estimated fit, standard errors, and number of observations on our reproduction all identically matched those reported by the WHR.

The out-of-sample error for this reproduction of the WHR model is 0.569. The out-of-sample error for the Dovecote model is 0.498.¹⁰

Given no evidence of over-fitting, a greater cross validation accuracy can be considered an improvement irrespective of the number of variables required to achieve it. However, in addition, it can also be demonstrated that the improvement is not purely a result of simply having more variables.

3.4.3 The WHR’s model fit can be surpassed entirely without WHR variables

To demonstrate this in one direction, it was found that it is possible to practically match the WHR’s 0.57 out-of-sample error with only *three* variables, and using not just the same data as the WHR itself, but with almost 250 *more* observations.

But perhaps even more striking, it can be demonstrated that *none* of the WHR variables are necessary to achieve a high model fit, in-sample or out-of-sample. The fit of its six-variable model can be surpassed using six *completely different* variables. Using only the very plausible set of variables on clean water, cities, pre-primary education, shelter, unemployment, gay and lesbian acceptance, and year fixed effects, an entirely different six-variable model achieved an adjusted R-squared of 0.76, identical to the WHR, and an out-of-sample error of 0.56, compared to the WHR’s 0.57. Both of these models can be seen in Table 3. Note that this model has *no* national economic measure.

Thus there is abundant material reason to doubt that the WHR’s hand-picked model specification should be treated as an exceptional representation of the phenomenon it describes – very much less to “explain” global life satisfaction, to use its own words.



Figure 2: Contributions of Models Equal to or Surpassing WHR Fit

¹⁰This is not a perfect comparison because, as discussed above, we do not use data from before 2011 in order to include the variable on gay and lesbian social acceptance. However, restricting the WHR model to our same time period produces an identical measurement of the WHR’s accuracy.

	Approximates WHR Fit	No WHR Vars
Social support	2.51*** (0.15)	
Feelings of freedom	1.93*** (0.10)	
log(GNI per capita)	0.42*** (0.01)	
Years of preprimary education .		0.08*** (0.01)
Shelter		0.01*** (0.00)
Gay + lesbian acceptance.		1.40*** (0.08)
Clean water		1.65*** (0.14)
Urban population		0.01*** (0.00)
Unemployment		-0.04*** (0.00)
Adj. R ²	0.74	0.76
CVE	0.59	0.56
Num. obs.	2199	1633

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 3: Two Small Models that Perform as Well or Better than the WHR’s Hand-Picked Model

3.4.4 Two out of six WHR variables are not chosen by the model in any form

Given this basic challenge to the six-variable specification, it is important to consider all of these six variables.

While the Dovecote model uses a different measure of lifespan and a different measure of the economy, it still uses close analogs to the versions of these two variables in the WHR model, and uses another two WHR variables directly. These four have been discussed, and their measured effects reported.

However there are two variables in the WHR model that have not yet been mentioned, because they are not considered to be used in any form in the Dovecote model; these are Generosity, and Perceptions of Corruption.

Generosity makes an almost negligible contribution to even the WHR’s model, changing the holdout error by less than a tenth of a point when it is removed completely. It makes a similarly insignificant and vanishingly small difference when added manually to the Dovecote model.

Perceptions of Corruption has a larger contribution when added to the Dovecote model, but still comes in 13th out of 16 in terms of contribution, does not meaningfully change the model fit, and is completely statistically insignificant.

While this is compelling evidence, manually inserting a variable does not demonstrate its relative importance in the large pool of possible variables. Experiments were run to produce estimates of the priority given to different variables in the search procedure. These test gradually varied the level of penalization, and then made note of the number of selected variables in the model when a variable was first selected in 100% of iterations. Using this method, Perceptions of Corruption and Generosity were respectively ranked at approximately the 23rd, and 28th most important variables – meaning, a model needed at least that many variables before it would choose to add them. As two variables out of just six, these constitute fully a third of the flagship WHR model.

One reason for this low ranking may be that perceptions of corruption are actually a relatively poor predictor

of government effectiveness, at least when using the definition provided by Varieties of Democracy (V-Dem). Unlike perceptions of corruption, V-Dem’s “government effectiveness” variable (`e_wbgi_gee`) is chosen in 100% of model iterations when using the data from 2005 onward. A plot of the V-Dem variable against the WHR variable shows only a very small portion of the support with a clear relation between the two, and an R^2 of just 0.25.

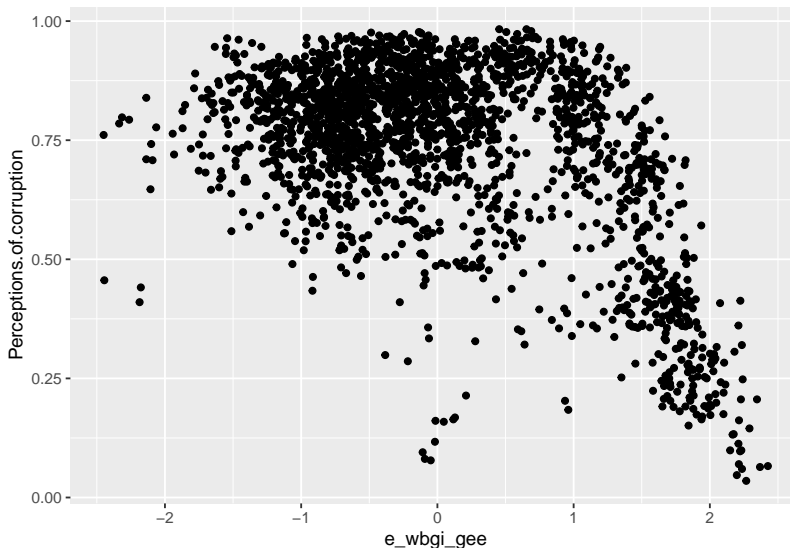


Figure 3: Government effectiveness vs. Perceptions of corruption

3.4.5 The effect of a logarithmic transformation on a national economy indicator

Before moving on to examine the Dovecote model in detail, it is necessary to explain one last major departure from the WHR’s specification. The WHR chooses to use the natural logarithm to transform the variable of GDP per capita. While we believe GNI to be a better choice, as mentioned above, the dramatic difference comes not from the choice of economy indicator, but from the use of the log transform.

The WHR claims its decision to use the transform is justified because it “fits the data significantly better than GDP per capita” (Helliwell et. al, 2023). However, not only do we find this to be at best marginally true in our reproduction of the hand-picked model (with the log lowering the holdout error from 0.58 to 0.57), we find it is definitely *not* the case in the more rigorously specified, and higher accuracy, Dovecote model.

To be precise, in the fourteen-variable, data-driven model, the adjusted R-squared actually goes down by .001, and the cross validation error goes up by .002, when GNI is replaced with log of GNI – a completely negligible difference, and actually a loss of model fit.

What is more, the distribution of residuals is visually indistinguishable, and in both cases very clearly not strongly skewed. Finally, the slight observed heteroscedasticity is also the same with or without the transformation. In summary, we find no mathematical rationale for the transformation of GNI or GDP at all.

3.4.6 Simple plots of the relationship are misleading

Some of the confusion may stem from the fact that a simple plot of satisfaction by log GNI does show a strongly logarithmic relationship. However, this two-way relationship is immaterial when considering the variables as part of a larger model. The real question is whether or not the log is an appropriate transform

for GNI when it is used as an input to regression. This can be easily checked by fitting a Generalized Additive Model (GAM) with the full set of variables from the Dovecote model, and applying a non-parametric transform just to the GNI term. This term will then have the flexibility to assume whatever form best fits the data. Figure 4 shows a plot of the non-parametric fit, overlaid in red with the best fit line, and the best fit logarithmic curve.

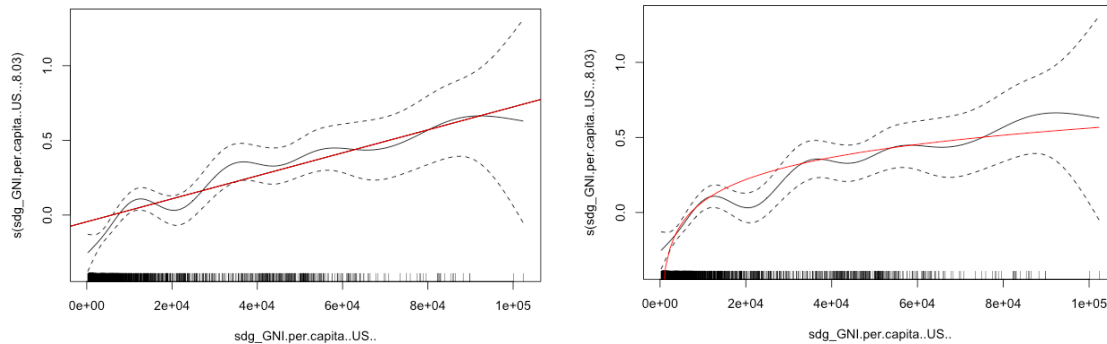


Figure 4: The Effect of a Log Transform on GDP

The line achieves an R^2 of 0.92, and the log achieves an R^2 of only 0.85. In other words, this is further evidence that, once all of the relevant variables are taken into account, the conditional relationship between satisfaction and GNI is best described linearly.¹¹

If, despite this superior fit, there is still a deep need to represent this relationship in a non-linear way, that can always be done with a piecewise-linear model. This would keep the parameters comparable, while completely avoiding the most serious problem with the log transform.

This basic disagreement with three out of six of the current model's parameters is also one reason we do not attempt to find variables to add to it. Especially given this empirical evidence, we strongly believe that starting from scratch is a better approach to a full and accurate specification.

3.4.7 The log drastically overstates the contribution of GDP

Were the log in fact to improve model fit, then there would be some justification for its use. However since it does not, then it is especially important to consider its multiple and severe disadvantages.

The first is that it is wholly unintuitive. While seasoned researchers may have developed some intuition for what it means to have a linear relationship with a logarithmic variable, this is certainly not the case for the public.

Even for those who do have such an understanding, the specific relationship is almost absurd. Increasing the value of log GDP per capita by 1 would require multiplying the un-transformed GDP per capita of by 2.7. Or in other words, the log-transformed model suggests that, were the relationship causal and precisely as estimated, then it would require almost tripling a country's GDP before achieving the points of increased satisfaction shown by the model. This must then be compared with numbers such as the increase suggested by *adding* one year of pre-primary education.

This basic incomparability is of course not unique to GDP, but the fundamental problem with interpretation when mixing transformed and un-transformed variables. One is describing an additive effect, and one is

¹¹All fourteen variables were assessed for non-linearity, and only the portion of the population living in cities appeared appropriate for a transform. The trend is sharply upwards from 0 to 85% of the population, but then sharply downwards from 85 to 100%. However the effect on the overall model is modest, and so the specification is kept entirely linear for parsimony.

describing a multiplicative effect. This makes them, at best, a challenge for interpretation and communication that requires substantial extra care.

The enormous problem with the transform in this context is not only that the WHR uses it, but that the report presents the fundamentally incomparable transformed results in a way that clearly suggests they are directly comparable with the transformed ones – and that this approach results in an enormous exaggeration of the effect.

Using the same 1,964 observations that are in the WHR’s reported model, Figure 5 shows the difference in estimated average variable contributions when using un-logged, vs logged, GDP in the WHR specification. With an un-logged variable, the contribution is measured at 0.30. With the logged variable, the contribution is measured at over 3.4 – more than eleven times greater. In the Dovecote model as well, the use of a log transform on only GNI appears to exaggerate its contribution by a factor of more than twelve.

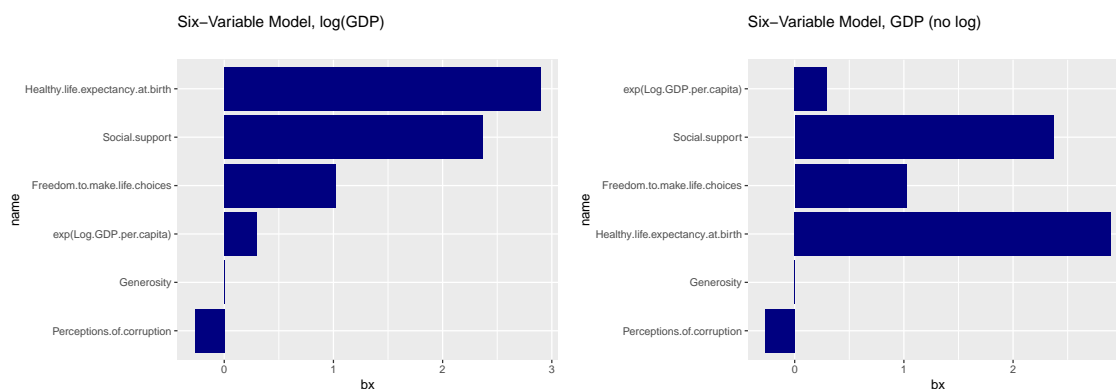


Figure 5: The Effect of a Log Transform on GDP

To reiterate for the sake of total clarity, this is not due to a more accurate representation of the data – it is because the log fundamentally changes the very meaning of the estimated “contribution.”

This presentation is an even more concerning issue when, rather than carefully accommodating the differences in meaning, the WHR repeatedly omits any mention of the transformation at all, and treats the model contributions as if they are directly comparable. They are not.

This direct comparison is most prominent in a multi-page graphic that the WHR has so far used in 2015, 2016, 2017, 2019, 2020, 2021, and 2022, where each surveyed country’s satisfaction is broken down into pieces corresponding to the model’s six variables. Log GDP per capita is simply labeled as “Explained by: GDP per capita,” and the logarithmic contribution is placed alongside the linear contributions. This presentation is not just irresponsible, but outright and materially misleading.

Far beyond a single graph though, implications of this seemingly minor technical detail stretch to the very core of the report’s message.

The WHR published in 2023 begins with an Executive Summary and a page of acknowledgments. The very next page after that is dedicated entirely to describing the hand-picked model, which it does in the following terms, here quoted in their entirety:

“Six Factors Explained

Income, health, having someone to count on, having a sense of freedom to make key life decisions, generosity, and the absence of corruption all play strong roles in supporting life evaluations.”

The page after that is dedicated entirely, without any mention of a logarithm, to “GDP per capita” – the very first of the six variables to be described. The page provides a brief and non-technical definition of GDP, but the majority of the space is devoted to a cartoon of people watering a plant bearing heart-shaped vegetables.

This presentation makes two things unambiguous. The first is that the WHR is targeting itself at a far broader audience than quantitative researchers.

The second is that no matter how many supplemental analyses they include, the WHR presents their hand-picked model as the cornerstone of the report.

The consequence of all of this is that a casual reader, and in fact even a well-trained reader, could easily come away believing they had been told that personal income is the single greatest factor in global happiness, and accounts on its own for almost 40% of model variation.

To address just one of two enormously misleading descriptions, we note that GDP per capita is both defined, and measured, as a completely different concept than personal or household income (Nolan, Roser, and Thewissen, 2016).

But more critically, the most accurate description we have been able to produce with this data says that the national economic indicator of GNI per capita contributes barely 2.3% to the model of average national satisfaction, as the 12th most important variable out of fourteen.¹²

3.4.8 The point of using life satisfaction is to avoid the mistakes of GDP – not repeat them

National economic indicators have been severely questioned for many years, and it is worth briefly reviewing some of the arguments against them.

Included among their shortcomings is that GDP and GNI have long failed to effectively account for environmental degradation or carbon emissions, which is especially concerning when they are so tightly correlated. Were these extraordinary externalities to properly accounted for, they may not just diminish GDP’s value, but in fact reverse it, based on the current structure of the economy.

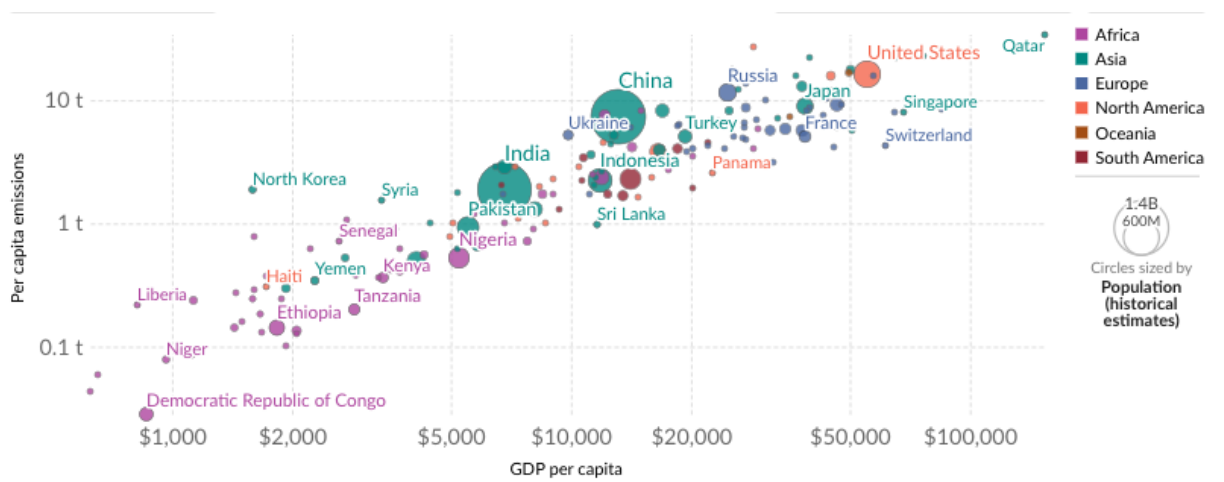


Figure 6: National Carbon Emissions by GDP

Others failings include the poor job that GDP and GNI do when estimating the quality of even basic and common resources such as education, technology, and medical care; and its even greater inability to accurately estimate the value of any service not performed for pay, including such survival fundamentals as parenting and child rearing.

¹²An argument in support of GDP could fairly point out that several of the variables in the Dovecote model are clearly factored into GDP in various ways, meaning there may be problems with overlapping definitions. However, given the extraordinarily vast range of activities that are accounted for in such a measure, simply clarifying which *parts* of GDP are predictive of satisfaction is an enormous increase in precision. This also avoids the easy but erroneous thinking that “GDP” predicts satisfaction, and so anything inside of the umbrella definition of “GDP” does, as well. Even as we believe intelligent economic processes to be critical in helping to supply the discovered variables, in terms of precision, actionability, policy relevance, and simple accuracy, we believe the Dovecote model still presents strict and substantial improvements over the current state of the art.

Social Resources		Material Resources	
Social Inclusion	Social Support Gay + Lesbian Acceptance Intimate Partner Violence	Health + Survival	Electricity Clean Water Life Expectancy
Self Determination	Feelings of Freedom No Shadow Governance	Monetary Economy	Unemployment GNI per capita
Inclusive Power	Women’s Political Power Public Campaign Financing	Education / Info	Years of Pre-primary Education
Cities		Cities	

Table 4: An Informal Grouping of the Dovecote Model Variables

Though for all the seriousness of these problems, it is perhaps most disturbing that for many people, the entire purpose of measuring life satisfaction is to *replace* GDP, due to these many, fundamental, well known and documented failures (Sen, Stiglitz, and Fitoussi, 2010).

So it is especially difficult to see this apparent 1100% exaggeration of GDP in the internationally distributed and globally respected “World Happiness Report.”

4 Discussion

4.1 The model variables

4.1.1 Categories of variables

Having addressed the largest differences with the WHR model, we now consider the Dovecote model in detail.

The fourteen variables included in our full model can be seen in Table 4. The first observation we make about these variables is that they provide strong evidence for the initial motivating suspicion – many critical variables do in fact appear to be missing from the WHR model.

For ease of consideration, we loosely group the variables into the two categories of Social Resources and Material Resources. Within these, we again loosely divide the variables into the seven subgroups of Social Inclusion, Self Determination, Inclusive Power, Education, Monetary Economy, Health and Survival, and Cities.

The number of people living in cities is placed in a category by itself, and included in both halves of the table, because it has clear contributions to both social and material resources. It is interpreted as primarily providing a more dense concentration of the other listed variables, though there are likely other important effects such as increased exposure to a diversity of people and information.

4.1.2 The contributions of the variables

Figure 7 shows the data-driven variables by subgroup, and the summed magnitudes of each subgroup’s contribution. The absolute value is used for variables with a negative contribution, so that the magnitude of all effects can be seen.

With this grouping, the categories of Social and Material Resources are relatively matched in total contribution, with Material Resources at 2.0 and Social Resources at 2.8, though caution should be used given the subjectivity of these groupings, and the complex nature of many of the individual variables.

This grouping is also used to suggest that, with further research, the specific variables currently within each category may well change. Our choice to include a category on “Education and Information” reflects our belief that it will remain an important component of any future model, even as we collect far more accurate and informative data on kinds of education, and types of information.

Though even now, this organization is helpful in seeing gaps in the model variables. In particular, this presentation illuminates one area that may be considered to have been structurally omitted: environmental sustainability. By “structurally omitted” we mean that a model of present satisfaction inherently ignores the difficult but critical question of future satisfaction. While we will not address this challenge formally here, it is worth noting that the findings of the Dovecote model – unlike the GDP-heavy WHR model – prioritize non-resource-intensive variables. Guiding policy according to its estimated priorities would likely have a strong positive impact on sustainability, even without intentionally taking it into account.

The last immediate issue we will identify with the selected variables is that the international availability of data does not appear to reflect the importance of that data. While incredibly detailed information is available on measures such as economic outputs, industry sectors, and international trade balances, only one type of such a variable – GDP/GNI – was actually selected by the model, and provided only a very marginal contribution to it. By contrast, the WHR’s variable on social support was essentially unique in its availability even among the 5,000 initial candidates, and yet it utterly dominates every single other variable in the model. The single variable on Social Support is estimated to contribute almost 30% of model variation, while an even distribution of effects over the 14 variables would predict a contribution of just 7%.

This qualitative result has independent replication as well, in studies no less than the almost century-old Harvard Study of Adult Development (Vaillant, 2015). If we know social relationships to be foundational, then the effort to collect data on them at scale should match their measured importance. If a category like “health” can meaningfully sustain international data collection on hundreds or thousands of individual diseases, conditions, and outcomes – a remarkable number of which appeared to independently support predictions of national satisfaction – then it will obviously be worth the effort to move beyond a very literally one-dimensional understanding of human relationships, as well.

Having examined the Dovecote model, the next important step is to understand what more is necessary before these findings can be used to guide policy.

4.1.3 Causal assumptions and policy relevance

From a strict econometric perspective, the variables in the Dovecote model are only predictive. There is no intervention or other instrument used to estimate a rigorous causal effect. However, the same is true of the WHR model, and yet it consistently states that its variables “explain” the national outcomes. This borders on causal language, and is naturally interpreted by any lay reader as causal.

While this is language that not everyone would find appropriate given the methods being used, it is true that the goal is to identify causal variables – and, it is valuable to consider the extraordinary difficulty of making traditional causal estimates on national averages. With even catastrophic events causing only small numerical shifts in an individual, it is difficult to imagine any simple procedure short of armed conflict that might measurably change an entire country’s satisfaction. What is more, any such event is certain to also affect the other variables in the model, making isolation and instrumentation exceptionally challenging. This rules out many approaches for causal estimation that might be employed in smaller systems.

Consequently, it is crucial to consider all other possible means for assessing causality. And in that context, we will point out that relative to the WHR model, the Dovecote model has several distinct advantages.

The first advantage is the extreme care that the data-driven Dovecote model has taken to fully explore the space of predictors. This means there is a smaller risk of capturing spurious, non-causal proxy variables, as there is a greater chance that the true causes were available to choose from. By contrast, the hand-picked WHR variables of Generosity and Perceptions of Corruption never appeared in any data-driven model before far more variables were allowed. As two variables of just six in the WHR specification, these constitute fully one third of the hand-picked model.

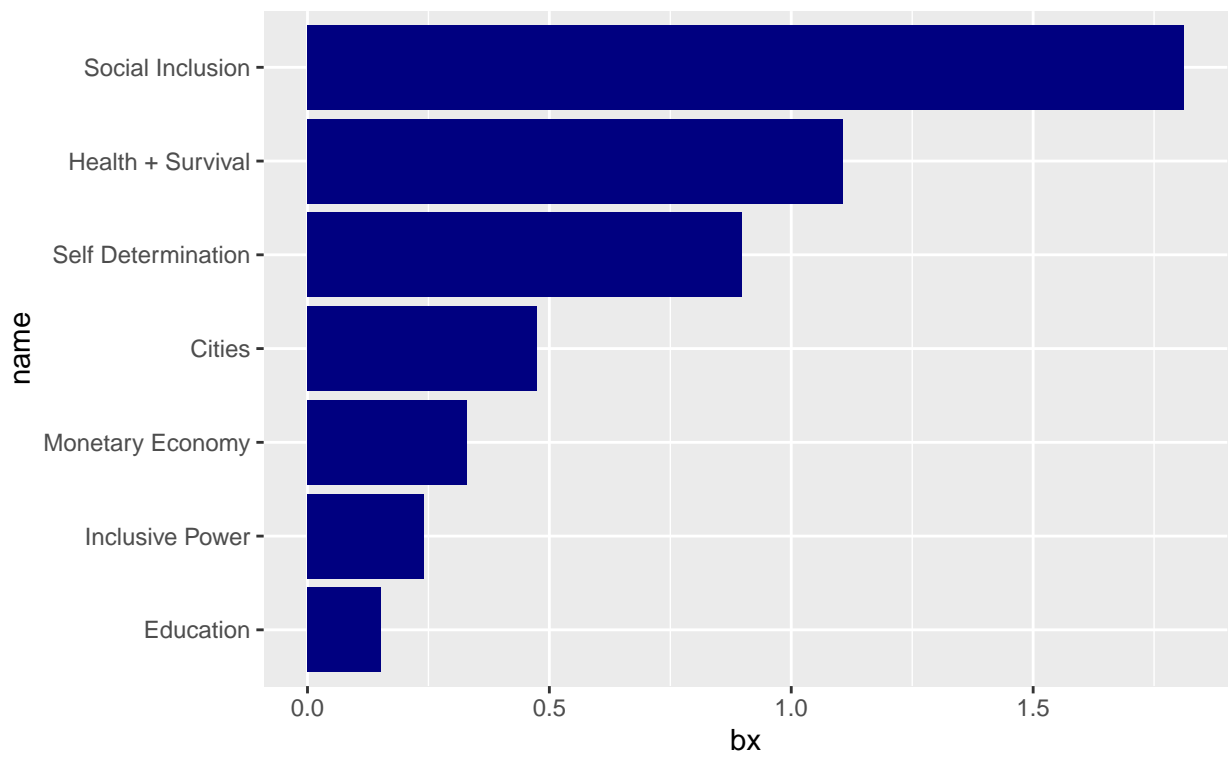


Figure 7: Contributions by Variable Category

The second advantage of the Dovecote model is the objective nature of the variables it selected. Variables like the WHR’s Perceptions of Corruption can make a strong argument for directly causing an individual’s satisfaction to change – this perception has a direct mechanism through the individual’s psychology. However from a policy perspective, the variable’s literal message is that simply hiding corruption will solve the problem. By contrast, the Dovecote model’s findings have clear and policy relevant implications. For example, they suggest increasing the availability of education, establishing “significant” public funding for political campaigns, and increasing the number of women in the legislature.

One major exception is the variable on Feelings of Freedom, which appears in both models. This is another self-reported variable, in which the respondents are asked the question: “Are you satisfied or dissatisfied with your freedom to choose what you do with your life?”

While this variable makes a clear and substantial contribution to predictions, in its current form, it can make only a very limited contribution to policy discussions. To increase its relevance, it is critical to better understand what circumstances, social, material, or otherwise, *produce* feelings of freedom – especially when in public discourse, “freedom” is frequently used as an effective argument *against* the other variables found in the model.

While acknowledging the correlative nature of the Dovecote model, given all relevant factors, we choose to make the assumption that the estimates for its variables are directionally causal.¹³

This is to say we assume that while the numerical estimates may not be precise, shocks to them would on average move national satisfaction in the estimated directions, either directly, or indirectly.

We strongly believe that substantial further work is necessary to test and clarify these relationships, before, during, and after the process of implementing policy based upon them – as well as to continue to look for improved data, better measurements, more variables, and the possibly specific conditions to which they pertain.

This caution is motivated in part by the fact that even among the discovered variables, multiple nuances are already known. For one example, women’s political power is almost certainly a partial proxy for the political power of other under-empowered groups – and that, when elected only in small numbers, it is altogether possible to elect women who make identical decisions as their male colleagues.

For another example, public funding for elections often goes to people who were intending to run anyway, and must be supplemented with targeted outreach efforts and social encouragement to run (Carnes, 2019). In other words, these variables may represent a very strong policy framework, but it would not be sufficient to attempt to implement many of them naively.

We believe furthermore that it is likely only possible to truly clarify a number of these variables and their effects when considered at the individual level, and not the national one – a recommendation already made by an international working group that reviewed alternatives to GDP.

As the data and research evolve, we fully expect the proposed model to be improved upon, clarified, and even changed in unexpected ways. In anticipation of this, we do not even claim that this specific model should definitively replace the WHR’s six-variable specification. Rather, we strongly recommend that a new model search should be done on an annual basis, and the flagship model continuously updated with our growing understanding. In every case, more research is enthusiastically encouraged.

However, in the current absence of any such process, we also believe that the risks of our proposed model, and of this directionally causal assumption, are outweighed by the clarity they provide to the consequent reasoning and conclusions.

We believe these risks to be small based not only on the specification process, but also the results: that life satisfaction at a national level is caused primarily by concrete and specific survival needs, meaningful social inclusion, actual control over your own fate, and an accurate grasp of reality.

¹³Though we add the important caveat that there can be substantial shocks to GNI per capita without equal movement to the median resources for an individual, and other critical statistics of the distribution (Sen, Stiglitz, and Fitoussi, 2010).

An inverse framing is also clarifying. Arguments that any of these do *not* cause satisfaction would rightfully inspire substantial skepticism. Certainly this should be the case for the WHR’s arguments, that it is possible to have a meaningful model without any reference at all to education, discrimination, infrastructure, or political power.

4.2 Governance, institutionalism, and paths towards the model variables

Considering causality is critical because the value of a model of life satisfaction lies in its ability not just to describe satisfaction, but to improve that satisfaction. Furthermore, to take this improvement seriously requires an understanding of not just parameter estimates and interpretation, but also how relevant interventions might take place. In turn, this means identifying the entities that have both the power and the inclination to make those interventions. However, the variables that the Dovecote model discovered in this area were very different from what was initially anticipated.

4.2.1 “Institutional” variables were available, but not selected

In looking for the paths through which to effectively intervene on the discovered variables, the initial expectation was for the model to return descriptions of formal institutions, or governance structures, such as a bicameral legislature – however, not one of these appeared.

This is clearly not due to a lack of opportunities. The V-Dem dataset makes available almost two thousand variables, including multiple scales, complex indices, and statistical moments for their many questions (Coppedge et. al, 2023).

Among these variables, some of which contain observations dating back to 1789, there are extensive collections on the structure and powers of the legislature, the head of state, the judiciary, elections, deliberation, the media, social exclusion, political equality, and civil and academic society, as well as many high-level indicators based on these components.

Specific variables include “Judicial constraints on the executive index (D)” (which is indexed under the code v2x_jucon), “Legislative constraints on the executive index (D)” (v2xlg_legcon), “Election vote buying (C)” (v2elvotbuy), “Social class equality in respect for civil liberty (C)” (v2clacjust), “Obligatory referendum participation threshold (A)” (v2ddpartor), and hundreds more.

4.2.2 The contributions come from availability, equality, and substance of political power

However, when all of these were made available as predictors of average national satisfaction, there were only three that survived in the final model. These are “Women political participation index (D)” (v2x_genpp), “Public campaign finance (C)” (v2elpubfn), and “HOS control over (C)” (v2exctlhs). As described above, these represent the amount of formal and informal political power held by women, the availability of “significant” amounts of public funding for political campaigns, and the absence of a body, such as the church or military, but most specifically from the ruling political party, from whom the head of state must seek approval before making major policy decisions.

While initially unexpected, these variables tell a clear and consistent story. They suggest that national satisfaction is dependent upon not just equitable and meaningful access to political power, but also upon a guarantee that this power will actually mean something once it is assumed.

From this perspective, it is easy to hypothesize that no other variables were selected because when these requirements are not fulfilled, the details of what happens downstream become inconsequential. Systems in which voting is technically allowed clearly cannot benefit when only un-elected individuals actually wield any power. Systems in which everyone is legally allowed to run for office, but the vast majority of people are not financially able to do so, recall the observation attributed to Mark Twain that “The man who does not read has no advantage over the man who cannot read.”

4.2.3 Published conversations continue to focus heavily elsewhere

These findings are especially surprising in part because current conversations in both the news media and academic literature tend to have a high focus on the role and functioning of the kind of institutions that the model did not select. As just one example, *Why Nations Fail* was a nationally best-selling book, written by well-regarded professors of economics (Acemoglu and Robinson, 2012). Its authors argue that the history of human civilization can largely be explained by “institutions” – frequently providing property rights, in particular, as an example of a key institution.

However, there were at least five separate versions of property rights variables among the 1,058 available to the model, describing property rights for private citizens, separate genders, and the state, and not one of them was ever selected, or in fact even sighted.

Property rights are of course only one kind of institution, and only one part of a larger argument. But if “institution” is instead meant in the sense of a bicameral legislature, or a particular organization of the court system, then once again, the model provides no support at all for any particular structure – with the possible exception of the inverted requirement that *no* un-elected groups, including appointed ones, should hold the powers of elected officials.

This almost total failure of the anticipated variables on governance structures may be surprising, but it is also robust, and should be taken seriously. Upon consideration, it also admits a very cohesive interpretation.

One fundamental problem is that whether “institution” is meant in either of the above ways, or in practically any other of the frequent uses, such as “respect for the judiciary” or “being polite on the floor of the House,” then it is still focusing on an intermediary abstraction that is hoped to provide *some* good outcome, but fails too often to effectively define that outcome. A related and equally deep problem is that without a quantifiable outcome, it is not possible to rigorously rank priorities. This makes it easy for policies with high emotional salience to be considered before much more impactful, but less exciting factors.

By contrast, when life satisfaction is used as a clear, measurable, explicitly and globally inclusive, conceptually rigorous outcome, then the model does not support complex processes, or patent systems, or property rights. It supports diversity in elected government, money for *putting* diversity in the government, and for that elected government to have real power – which upon reflection appears almost self-evident. An unavoidable prerequisite of widespread satisfaction is reasonable equality of effective political power, and providing guaranteed resources specifically towards that goal is unavoidably the single way to reliably achieve it.¹⁴

In this finding, we are in broad qualitative agreement with *Why Nations Fail*, which argues that institutions must be “inclusive” and not “extractive” to ensure a nation’s success. This argument is strong, and amply supported. However within these extremely broad guidelines, we find no specific evidence that institutions such as property rights or patent systems rank anywhere near the top of the list for a nation that finds life worthwhile.

Perhaps even more importantly, a total focus on “institutions” runs the risk of relinquishing humans of any role in their own circumstances, which is a practical problem much more than a philosophical one. Not only is an “institution” an abstraction, not technically able to make any decisions on its own, but focusing on these abstractions tends to encourage similarly abstract theories about their reform.

For example, *Why Nations Fail* suggests that large institutional changes are only possible during “critical junctures,” which are massive and exogenous historical events such as Europe first traveling to the Americas, or the Black Death in the Middle Ages. While this may be helpful in retrospective analysis, we have no desire, or intention, to depend upon the plague as a prerequisite for improving the modern political system.

But where do we then turn?

¹⁴It is worth noting that the variable on public funding for political campaigns is small relative to other contributions, but shockingly specific and actionable. Only it, and years of education, are precise enough to even be considered policies. Variables like “social support” can be greatly influenced by law, such as on de jure and de facto segregation, but the changes themselves are complex, diverse, cannot be directly legislated into being, and would take years to fully materialize. By contrast, and uniquely, public funding for campaigns could in theory be voted into existence tomorrow.

4.2.4 The characteristics of individuals as components of a political system

One of the possible reasons that more institutions were not selected by the model is that institutions have limited structural validity. They are frequently composed of hundreds or even thousands of changeable and shifting rules, laws, norms, and circumstances. If we are to find a solid foundation for national-scale interventions, then it is likely necessary to focus on micro-foundations, or the atomic components of institutions. Thus, we choose to focus on the single existing component of any system that is able to directly make decisions and take actions – the individual.

Perhaps the most fundamental argument for focusing on individuals comes from considering the simple nature of institutions. There is no doubt that institutions of some kind are critical, inevitable, and even fundamental to any large group of humans, extending far beyond formal systems of governance. But even at its very best, every institution at its core is no more than a codification of behavior. Scholars, governments, and citizens around the world now describe our most treasured institutions as under attack, or failing, or already failed. But by this dramatic description, they mean nothing more complicated than that the people in these institutions are not demonstrating the behaviors we hope for.

We can design new incentives, and devise more rules. But when the only people in the world able to codify these new rules are the same ones who are already refusing to act upon the old ones, there is only a single avenue that remains. By all considerations, legal, theoretical, and practical, the sole possible solution is to be certain of the behaviors we truly need, and give power to the individuals who can actually be trusted to perform them. There can be no institutions without people – and thus there can be no institutional reform without people who have the power and will to choose reform. Our task then inevitably becomes to identify such people, and invest them with the necessary powers.

For several reasons, this can be particularly challenging. From an academic perspective, not only are politicians much less likely to respond to surveys, especially when they have politically sensitive questions about beliefs and intentions, but the relatively small impact of most individuals makes it very hard to connect a personality type to a national outcome.

However, this is not the only path for inference. Six variables can paint only a very limited picture, but the detailed set of fourteen variables discovered by the Dovecote model are able to be far more descriptive. These variables display a pattern that is highly consistent with two well-known and well-studied psychological traits.

4.3 Social dominance orientation and right wing authoritarianism

These traits are known as Social Dominance Orientation (SDO) (Sidanius and Pratto, 2001) and Right Wing Authoritarianism (RWA) (Altemeyer, 2008).

SDO and RWA are two major theories from the generalized prejudice literature in psychology – ideas attempting to explain not just a single behavior such as wage gaps or lynch mobs, but rather the general phenomenon of out-group discrimination. These theories have seen a resurgence of interest in recent years as they have proven to be exceptionally strong predictors of contemporary political extremism and voter behavior, most notably around the monumental shifts of Brexit (de Zavala, Guerra, and Simao, 2017) and the election of Donald Trump (Womick et. al, 2019).

SDO describes the extent to which a person believes in a group-based social hierarchy. It correlates highly with racism, sexism, homophobia, and other similar group-based prejudices. A person high in SDO believes in, and acts on, social Darwinism. They spend substantial effort building power for themselves, often at the expense of others.

RWA describes the extent to which a person submits to, and enforces the mandates of, established authority figures, as well as their tendency for social and cultural conventionalism. A defining consequence of these behaviors is that a person high in RWA will believe and obey the chosen authority despite any conflict in logic, evidence, or sometimes even visible reality. RWA is also frequently associated with an extreme degree

of information control, which can be enforced both internally and externally. Some high-RWA religious communities consider it a sin to even be exposed to information that might test a follower's faith.

One critical finding from these literatures is that, while both represent a high degree of out-group prejudice, SDO and RWA are largely uncorrelated. One represents a desire to dominate, and the other a desire to fit in. However uncorrelated does not mean anti-correlated. It is entirely possible to have both in high levels in the same individual – and when these traits appear simultaneously, they produce an exceptionally dangerous combination of behaviors.

These individuals high on both scales believe in a strong social hierarchy, and that they belong at its peak. They fight for power no matter the legitimacy of their own claim to it, easily justify immoral behavior, reject all undesirable information, and surpass people high in only RWA or SDO on measures of xenophobia. Unlike individuals high only in RWA, who are primarily followers, these “double highs” (Altemeyer, 2008) are authoritarian leaders. This makes them extremely effective at mobilizing high-RWA followers, as they fit exactly the desired authoritarian archetype.

They are pertinent in this context because together, these two personality traits have direct and immediate relevance to fully ten of the fourteen discovered variables.

4.3.1 Social support

Social support cannot be relied upon from individuals high in SDO, which follows a law-of-the-jungle style of individualism. While individuals are not completely anti-social, they exhibit a high favoritism for only those who they perceive to be high-status (Levin et. al, 2002).

4.3.2 Gay and lesbian acceptance

Gay and lesbian acceptance conflicts with SDO through the fact that LGBTQ+ individuals are frequently discriminated against, meaning they are often considered lower status (Poushter and Kent, 2020). Acceptance conflicts more consistently with RWA, through RWA's strong bias towards cultural conventionalism (Crawford et. al, 2015).

4.3.3 Feelings of freedom

Feelings of freedom conflicts strongly with RWA, as this freedom is necessarily at the expense of the control of the chosen authority. Robert Altemeyer, who developed the theory of RWA and remained its chief expert for decades, went so far as to title one of his books on the subject *Enemies of freedom: Understanding right-wing authoritarianism*. (Altemeyer, 1988).

4.3.4 Shadow government

Not having a shadow government conflicts strongly with SDO, where individuals strive to take control under all circumstances. In a series of experiments, participants high in SDO voluntarily assumed power at rates two times higher than individuals low in SDO, and four times higher than authoritarian followers, who are individuals high in only RWA (Altemeyer, 2008). The high SDO participants also effectively usurped power from the formally established authorities when power was not given to them through formal What's more, attempting to create an un-elected government by capturing the judiciary is a standard and well-recognized tactic of authoritarians, with a known and terrifying effectiveness (Ziblatt and Levitsky, 2018).

4.3.5 Inclusive power

The category of Inclusive Power conflicts with the behaviors of both SDO and RWA. Both of its variables, political power for women and public funding for election campaigns, democratize power in a way that is directly threatening to both established authorities and hierarchies (Austin and Jackson, 2019), (Gomez and Doherty, 2021).

4.3.6 Education

Both of the variables in the category of Education, whether the model uses pre-primary or compulsory, conflict with the extreme information control inherent to RWA systems such as fundamentalist religions (Altemeyer, 2008). Even more directly, a successful education also conflicts with the behavior described by research findings that individuals high in RWA are less likely to update their beliefs when presented with conflicting information (Sinclair, Stanley, and Seli, 2020).

4.3.7 Unemployment

The variable on unemployment represents a group of people with limited amounts of power. SDO is defined by a belief that such power differentials are morally right, and should be maintained. These beliefs are consistent even for circumstances over which the sufferer has no control, such as mental illness (Bizer, Hart, and Jekogian, 2012).

4.3.8 Cities / the proportion of the urban population

While cities are inherently complex, and expose inhabitants to an enormous number of different variables, one of the most distinct is racial, social, class, and other diversity. Studies of college students have shown that the RWA level of high-RWA freshmen is substantially reduced with extended exposure to higher education (Altemeyer, 2008). The effect is hypothesized to be a result of access, possibly for the first time, to people and information that do not fit into previous, highly curated, experiences, attenuating the harsh and unfounded criticism of these previously foreign groups. Cities are hypothesized to have a similar effect, through a similar mechanism. It also is notable that cities are frequently denounced as immoral, dangerous, and taboo by high RWA communities (Dawson, J, 2011).

4.3.9 Water, electricity, lifespan, and GNI

Only the variables on such fundamental needs as clean water, electricity, life expectancy, and GNI do not have obvious immediate connections to RWA and SDO. However, the combination of the two traits produces a profile that is strongly resistant to undesirable facts, and without any interest in actions that do not produce personal benefit. In resource management studies, the worst outcomes were reliably produced by teams with these “double high” individuals (Altemeyer, 2003), strongly suggesting this is not a profile that can be expected to carefully or effectively manage even such basic policy areas.

4.4 SDO, RWA, and the psychology of decision makers

Having performed an extensive variable search for the drivers of satisfaction, we discovered a set of highly plausible candidates for intervention, and then began a search for the paths through which these interventions might be made. With almost no institutional guidance from the model, we shifted our focus to the role of individuals. With that new focus, it was possible to see extensive connections between the fourteen model variables and the personality traits of social dominance orientation and right wing authoritarianism. The final connecting argument now requires us to demonstrate that empowering individuals low in both of these traits is the most likely path to increase national satisfaction.

In particular, we focus specifically on government decisions makers. This is both because many of the variables in the model are explicitly governmental, but also because government is the only sector that admits interventions for its most important decision makers. We cannot elect people to be rich, or famous, or the heads of major enterprises, even if people in these positions may be critical to the final outcome. But we can elect people to lead the government, and to design the systems in which all of the other participants act – so that is where we focus our attention.

In order to avoid old and often dishearteningly religious debates, we intentionally and entirely avoid the question of whether these decision makers would prefer or enforce any particular school of governance – we focus only on the question of how they would effect national life satisfaction, and the variables believed to produce that satisfaction.

It is important to note that it may only ever be indirectly possible to demonstrate a clean causal link between national satisfaction and the psychology of government leaders. This is because, given the extraordinary scale and flexibility the problem, it may never be possible to produce comparable outcome-based evidence for a set of desired psychological traits in the small handful of people whose decisions have national consequences. At a minimum, it may never be possible with the kind of sample size and control that might be desired.

However, even if the evidence is indirect, that does not mean it is weak. In fact, some of the most direct evidence is also rigorously causal.

Experiments using the Global Change Game, a model-UN-like simulation, selected participants based on their tested SDO and RWA scores (Altemeyer, 2003). A round of the game designed to have all high-RWA players had multiple billion deaths simply due to the player’s almost total inaction, as the authoritarian-follower personality waited for somebody else to solve the problems. A cohort with all high-RWA, but also a small number of high-SDO players (making them “double highs,” or authoritarian leaders), destroyed the entire population of the globe through a nuclear apocalypse – and then was about to do it again, when the study observers ended the game.¹⁵

Just as critically, there is not only strong evidence against double highs, but also clear evidence in favor of double lows. The round of the Global Change Game with only players low in RWA solved more problems, saved more lives, and generally cooperated with one another more, and more effectively, than any randomly selected team that the game designers had ever seen.¹⁶

Neither is there any shortage of suggestive historical evidence. Abraham Lincoln, as an individual, did not seek to dominate. He respected his political detractors so much he often put them in his own cabinet (Goodwin, 2005). The outcomes of Lincoln’s policy interventions are likely to be, and remain, the single greatest increase in life satisfaction the United States will ever see.

Mike Mansfield, who presided over what is now referred to in almost saintly terms as “Mansfield’s Senate,” was a soft-spoken, thoughtful, and inclusive leader, who cultivated a spirit of bipartisan cooperation and legislative productivity that has not been matched for decades (Ornstein and Mann, 2008).

These two examples are also valuable for repelling the common but incorrect assumption that a lack of aggression is equivalent to a lack of resolve. Lincoln, despite his personal meekness, was a determined, effective, and ultimately victorious Commander in Chief during four years of civil war (Goodwin, 2005). Mansfield was the man who calmly rebuffed Lyndon Johnson’s attempt to preside over the Senate while he was Vice President (Caro, 2002), which would have been an unprecedented assault on the separation of powers. It is not these kinds of decision makers but rather people high in RWA who are most likely to follow bad orders, when they are pressured by someone they consider to be an authority.

Still, a simple theoretical argument may be more stark than any example. Simply using an unweighted average of national satisfaction as an outcome takes an inherent stance that all people’s lives are considered seriously, whereas SDO and RWA are both defined by the innate premise that large groups of people are inherently less deserving. It is difficult to imagine a less reconcilable conflict.

¹⁵An immaculate example of why the authoritarian-leader and authoritarian-follower team is referred to as the “lethal union.”

¹⁶This particular game did not measure SDO, but is suspected to have had several high SDO players. These were the participants who repeatedly tried to invade other countries, but were stopped by their low-SDO colleagues.

4.4.1 Finding and running candidates low in SDO and RWA

With additional time and study there will certainly be evidence connecting additional traits to national life satisfaction, and nuances revealed in the present set. But these two literatures are also already decades old. Searches for the terms “right wing authoritarianism” and “social dominance orientation” each return over twenty thousand papers on Google Scholar (Google Scholar, 2023). These tens of thousands of study provide a very substantial amount of clarity already.

One finding of the literature is that both of these traits are highly stable over periods spanning at least ten years, further cementing their suitability for identifying individuals with particular profiles (Osborne et. al, 2020).

Thus by highlighting the personality constructs of SDO and RWA, the Dovecote model provides clear guidelines for action in support of a well-defined goal. These characteristics directly inform the kinds of candidates we must be able to identify, field, and support, in local and national political races. Whereas attempting to preserve institutions by looking for someone who respects, for example, the “institution of the Senate” runs a large and fundamental risk of missing the point entirely.

Still, large and difficult questions remain: how can individuals with these characteristics be reliably identified, especially when the current instruments for their evaluation rely on answers from the respondent, instead of an external evaluator? Once identified, how can these individuals be most effectively persuaded, and then supported, in running for political office? And of course, what other traits are critical to either seek or avoid, for their impact on population satisfaction – especially when these individuals must compete for their aims in political environments filled with powerful and established antagonists, who are likely to have the opposite psychological traits? These are the questions we are most inclined to pursue in future research, and experiments – though other efforts are enthusiastically encouraged, both within and completely outside of academia.

Should this approach feel contrived, it is valuable to compare it with the many identical efforts that already exist, but for very different ends. Relatively little is currently known about what recruiters search for, but available evidence suggests candidate recruiters typically focus on individuals who are known quantities with known positions, such as corporate lawyers, who are similar to the recruiters themselves, or to previously successful candidates (who are themselves most likely have been chosen with established criteria). Recruiters also have a measurable preference for white collar candidates, given a strong but incorrect belief that blue collar candidates are less likely to win, and frequently target individuals who are independently wealthy because that income makes it much easier to devote time to the race (Carnes, 2019).

It should be clear that *none* of these characteristics currently used in the dominant paradigm are reliable predictors of low SDO or low RWA – and in some respects, seem far more likely to produce *high* levels of both. This is critical because in surveys, most respondents make clear that being recruited was a critical or necessary step in their decision to actually run for office. Thus without better guidance and clear goals, the current approach to candidate recruitment could continue to deeply shape the character of national government in a way that is at best unrelated to, and at worst terribly destructive for, positive national outcomes.

Given the severity of the situation, we include one final entreaty, should this approach still feel inorganic, supported by insufficiently econometric evidence, or counter to the theoretical ideal of eager citizens meritocratically competing for civil servitude in good faith.

It takes little more than a glance at the current population of elected, appointed, and actively governing individuals to see the truly stupendous gulf between the leadership we require, and what is now terrifyingly common behavior. American approval of Congress has now hovered around 20% nationally for almost two decades (Gallup, 2023). Even after a mob attempt on the lives of their colleagues, 147 elected members of Congress voted to invalidate the results of the 2020 presidential election, an action tantamount to an attempted internal coup (Yourish, Buchanan, and Liu, 2021). And in the states, where things may be even worse, over a third of voter-passed citizen initiatives were changed or simply ignored by the local lawmakers (Whitesell, 2024). With the final advent of Donald Trump’s long-predicted success in the primaries, a man who attempted to violently overthrow the United States government is now in line to win the presidency.

Our current situation is, from any possible perspective, an outright crisis. It threatens not just our institutions, or our norms, or our standing in the world – but the very nature of our lives. We present the Dovecote model and its attendant reasoning as a guide that may be imperfect and incomplete, but adamant that these are not failings that keep from it the possibility of still being a spectacular improvement over the current state of politics. And, we argue, quite possibly the single viable path towards a better, democratic, and even simply functional system.

4.4.2 Democracy is unavoidably necessary – but it is also just the beginning

It is critical to emphasize that to talk about individuals does not mean that any individual should replace a system of government. The findings from the Dovecote model are unambiguously on the side of democracy.

While democracy is often thought of as its rules and processes, the model highlights the intended conclusions of those processes: broad and meaningful access to political power. Such an outcome is self-evidently impossible without processes such as universal suffrage, and just because a variable on elections being accepted was not selected (v2elaccept, “Election losers accept results (C)”) does not mean it is not part of the larger important picture. But as the model, and the raw V-Dem data make clear, these institutions are by no means sufficient. And with such limited time and resources, we must scrupulously prioritize our efforts.

Perhaps just as importantly, the public has made clear they are rapidly losing interest in the very concept of institutions.

For many good reasons, “democracy” is likely the institution that currently appears most frequently in national media – most often to warn of its continuing and precipitous decline (Williamson, 2023), (Leonhardt, 2022), (Klaas, 2022), (Sundaresan and Isackson, 2021), (Merritt, 2022). However, no matter how earnest those discussions, or ample their evidence, only 48% of polled Americans reported they “Agree strongly” that democracy is the best form of government (Helmstetter and Fraser, 2022), and only 7% of polled voters said fixing democracy was the most important problem facing the country (NYTimes/Siena Poll, 2022). In the face of this extraordinary apathy, responding with an argument reiterating democracy’s value is an emotionally understandable response – but a more effective one is likely to be talking about something that the country actually cares about.

To a population angry, fearful, frequently misinformed, often knowingly misled, and long since sick of abstractions without results, there will only ever be one reliable source of hope: a better life. Life satisfaction is a goal that is innately capable of uniting a divided public, because it is a true goal, to which the processes we hypothesize will always be secondary. So it is now a truly existential challenge for us to be able to honestly seek, and accurately identify, the ways to achieve those better lives. We risk losing the forest for the trees if discussion of productivity, or precedent, or parliamentary procedure overtakes action on meaningful movement towards a better world. We risk alienating the people we most need to engage, when we talk about the role of abstractions, while ignoring the irreplaceable role of private citizens. And we misdirect our hopes if we fail to consider the most fundamental component of any of the systems we are trying save: individuals making choices about other individuals. Let us do everything we can to find the individuals who will make those choices wisely.

5 Conclusion

Contemporary models of life satisfaction are, conservatively, the culmination of decades of research. To be more grandiose, if still defensible, they are an accomplishment wholly new to humanity – a quantitative, data-backed, truly empirical, guide to the lives that the population of the full globe finds meaningful.

However, over a decade after the academic manifesto “Beyond Money” (Diener and Seligman, 2004), written by two of the most significant figures in the social sciences, the generational dream of life satisfaction researchers has in some ways succeeded wildly, and in other ways badly stalled. Detailed global life satisfaction

data exists – but when talking to Gallup, the kind of customer they describe as typical is not an academic or policy researcher, but an insurance company trying to predict violent revolution so that they can incorporate it into their pricing. And when talking to academics, they make it clear that this is because of a prohibitive cost for that critical data.

There is some institutional acceptance of the idea of measurable life satisfaction – but a vanishing number of non-experts are even vaguely aware of its meaning, importance, or even existence.

There is an internationally distributed model of the drivers of satisfaction, and that model is taught in universities, and published in economics textbooks. But that model erroneously enshrines exactly the variable it was designed to replace, leading even the people who teach it to dismiss its conclusions.

Simultaneously, humanity is faced with a global resurgence of anti-democratic governments (“The global state of democracy 2023,” 2023), the humanitarian crises those governments reliably create (Flagg and Calderon, 2020), and innumerable environmental cataclysms. The Intergovernmental Panel on Climate Change (IPCC) estimates only seven years before we reach a “point of no return” for the Earth’s climate, and the current levels of atmospheric carbon have not been reached since the Jurassic period, when sea levels were almost 300 feet higher (Core Writing Team, H. Lee and J. Romero (eds.), 2023), (Wallace-Wells, 2019). We have now already surpassed a global increase of 1.5C, a threshold that the world’s experts have long referred to as the final threshold before historically unprecedented consequences.

At a point in history as volatile as this, and with already so many loud warnings about the dangers of authoritarianism, it might seem redundant, or even perverse, to try to solve humanity’s problems with regression. But in times of such extraordinary crisis and complexity, we stand our best, and perhaps only chance, by having clarity about our path out. A clear picture of where we must go, and how we will get there, is as urgent as it has ever been.

This model hopes to provide some of that clarity.

If we want a real chance at the lives to which we aspire, then we cannot be led by people high in right wing authoritarianism, and we cannot be led by people high in social dominance orientation – and above all, we cannot afford to be led by people who are high in both.

So while democratic political power remains available, it must be used to its fullest extent. We must choose the people who will help us achieve the lives to which we aspire – and reject those who we can already predict will not.

5.0.1 Acknowledgments

The author would like to thank Shahzoor Safdar for his excellent advice on multiple early drafts, and Jeremy Rosen for repeatedly helpful discussions and suggestions. Last, a particular thank you to John Helliwell and Jan-Emanuel De Neve – for very kindly correcting errors in the data and confidence intervals, for engaging in such a collegial manner with this direct challenge to their work, and of course, for constructing the base upon which this paper hopes to build.

6 Bibliography

- Acemoglu and Robinson. 2012. *Why nations fail: The origins of power, prosperity, and poverty*. New York: Crown Publishers.
- Altemeyer. 1988. *Enemies of freedom: Understanding right-wing authoritarianism*. <https://psycnet.apa.org/record/1988-98419-000>.
- Altemeyer. 2003. What happen happens when authoritarians inherit the earth? A simulation. *Analyses of Social Issues and Public Policy*, 3: 161–169.
- Altemeyer. 2008. *The authoritarians*. <https://theauthoritarians.org/>.

- Austin and Jackson. 2019. Benevolent and hostile sexism differentially predicted by facets of right-wing authoritarianism and social dominance orientation. *Personality and Individual Differences*. <https://www.sciencedirect.com/science/article/abs/pii/S0191886918305853>.
- Bizer, Hart, and Jekogian. 2012. *Belief in a just world and social dominance orientation: Evidence for a mediational pathway predicting negative attitudes and discrimination against individuals with mental illness*. <https://www.sciencedirect.com/science/article/abs/pii/S0191886911005095>.
- Cantril. 1965. *The pattern of human concerns*. Rutgers University Press.
- Carnes. 2019. *The cash ceiling: Why only the rich run for office and what we can do about it*.
- Caro. 2002. *Master of the senate: The years of lyndon johnson*.
- Coppedge et. al. 2023. *V-dem 2013-2023 dataset v13*. Varieties of Democracy (V-Dem) Project. <https://doi.org/10.23696/vdemds23>.
- Core Writing Team, H. Lee and J. Romero (eds.). 2023. *IPCC, 2023: Climate change 2023: Synthesis report. Contribution of working groups I, II and III to the sixth assessment report of the intergovernmental panel on climate change*. Geneva, Switzerland. doi.org/10.59327/IPCC/AR6-9789291691647.
- Crawford et. al. 2015. Right-wing authoritarianism predicts prejudice equally toward "gay men and lesbians" and "homosexuals". *J Pers Soc Psychol*. <https://pubmed.ncbi.nlm.nih.gov/26524000/>.
- Dawson, J. 2011. Taking our cities for god. *Last Days Ministries*. https://www.lastdaysministries.org/Articles/1000008613/Last_Days_Ministries/LDM/Discipleship_Teachings/John_Dawson/Taking_Our_Cities.aspx.
- de Zavala, Guerra, and Simao. 2017. *The relationship between the brexit vote and individual predictors of prejudice: Collective narcissism, right wing authoritarianism, social dominance orientation*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5712068/>.
- Diener and Seligman. 2004. *Beyond money: Toward an economy of well-being. Psychological science in the public interest*. <https://doi.org/10.1111/j.0963-7214.2004.00501001.x>.
- Diener et. al. 1985. The satisfaction with life scale. *J Pers Assess.*, 49(1): 71–75.
- Diener et. al. 2009. *Well-being for public policy*. Oxford University Press. <https://books.google.com/books?id=k3c9VDD6IPQC>.
- Flagg and Calderon. 2020. *500,000 kids, 30 million hours: Trump's vast expansion of child detention*. <https://www.themarshallproject.org/2020/10/30/500-000-kids-30-million-hours-trump-s-vast-expansion-of-child-detention>.
- Gallup. 2023. *Congress and the public*. <https://news.gallup.com/poll/1600/congress-public.aspx>.
- Gilman. 2023.
- Gomez and Doherty. 2021. *Pew Research Center*. <https://www.pewresearch.org/short-reads/2021/07/22/wide-partisan-divide-on-whether-voting-is-a-fundamental-right-or-a-privilege-with-responsibilities/>.
- Goodwin. 2005. *Team of rivals: The political genius of abraham lincoln*. Simon; Schuster.
- Google Scholar. 2023. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C9&q=%22right+wing+authoritarianism%22&btnG=.
- Helliwell et. al. 2023. *World happiness report 2023 (11th ed.)*. Sustainable Development Solutions Network.
- Helmstetter and Fraser. 2022. *CAPM research lab analysis of McCourtney institute's mood of the nation poll*. <https://www.apmresearchlab.org/motn/poll-americans-belief-in-democracy>.
- Infurna et. al. 2020. Changes in life satisfaction when losing one's spouse: Individual differences in anticipation, reaction, adaptation and longevity in the german socio-economic panel study (SOEP). *Aging and Society*, 37: 899–934.
- Klaas. 2022. *America's self-obsession is killing its democracy*. <https://www.theatlantic.com/ideas/archive/2022/07/american-democracy-breakdown-authoritarianism-rise/670580/>.
- Larrain. 2020. *Macroeconomics: MIT press essential knowledge series*. Gildan Media.
- Leonhardt. 2022. "A crisis coming": *The twin threats to american democracy*. <https://www.nytimes.com/2022/09/17/us/american-democracy-threats.html>.
- Levin et. al. 2002. *Social dominance orientation and intergroup bias: The legitimization of favoritism for high-status groups*. <https://journals.sagepub.com/doi/10.1177/0146167202282002>.
- Merritt. 2022. *Op-ed: Democracy's survival depends on fighting demagogues. The GOP is embracing them*. <https://www.latimes.com/opinion/story/2022-08-09/demagogues-democracy->

- donald-trump-republican-party-gatekeeper.
- Nolan, Roser, and Thewissen. 2016. GDP per capita versus median household income: What gives rise to divergence over time? *Social Macroeconomics Working Paper Series*. <https://www.bsg.ox.ac.uk/sites/default/files/2019-06/SM-WP-2016-05%20GDP%20per%20capita%20vs%20media%20household%20income.pdf>.
- Nolan, Roser, and Thewissen. 2016. *Social Macroeconomics, Working Paper Series*. <https://www.bsg.ox.ac.uk/sites/default/files/2019-06/SM-WP-2016-05%20GDP%20per%20capita%20vs%20media%20household%20income.pdf>.
- NYTimes/Siena Poll. 2022. *Cross-tabs for october 2022 times/siena poll of registered voters*. <https://www.nytimes.com/interactive/2022/10/18/upshot/times-siena-poll-registered-voters-crosstabs.html>.
- Ornstein and Mann. 2008. *The broken branch: How congress is failing american and how to get it back on track*. Oxford University Press.
- Osborne et. al. 2020. *Authoritarianism and social dominance predict annual increases in generalized prejudice*. <https://journals.sagepub.com/doi/abs/10.1177/1948550620969608>.
- Poushter and Kent. 2020. The global divide on homosexuality persists. *Pew Research Center*. <https://www.pewresearch.org/global/2020/06/25/global-divide-on-homosexuality-persists/>.
- SDG. 2023.
- Sen, Stiglitz, and Fitoussi. 2010. *Mis-measuring our lives : why GDP doesn't add up?* Post-Print no. hal-03415632, HAL. <https://ideas.repec.org/p/hal/journal/hal-03415632.html>.
- Sidanius and Pratto. 2001. *Social dominance orientation: : An intergroup theory of social hierarchy and oppression*. Cambridge University Press.
- Sinclair, Stanley, and Seli. 2020. Closed-minded cognition: Right-wing authoritarianism is negatively related to belief updating following prediction error. *Psychon Bull Rev*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7384563/>.
- SPI. 2023.
- Sundaresan and Isackson. 2021. *Democracy is declining in the u.s. But it's not all bad news, a report finds*. <https://www.npr.org/2021/12/01/1059896434/united-states-backsliding-democracy-donald-trump-january-6-capitol-attack>.
- The global state of democracy 2023: The new checks and balances*. 2023. <https://doi.org/10.31752/idea.2023.78>.
- Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B)*, 58: 267–288.
- Vaillant. 2015. Triumphs of experience: The men of the harvard grant study. *Belknap Press*.
- Wallace-Wells. 2019. *The uninhabitable earth*. Penguin Books.
- Whitesell. 2024. States legislatures, governors are increasingly undermining ballot measures that win. *Talking Points Memo*. <https://talkingpointsmemo.com/cafe/states-legislatures-governors-are-increasingly-undermining-ballot-measures-that-win>.
- WHR/Gallup. 2023.
- Williamson. 2023. *Understanding democratic decline in the united states*. <https://www.brookings.edu/articles/understanding-democratic-decline-in-the-united-states/>.
- Womick et. al. 2019. Group-based dominance and authoritarian aggression predict support for donald trump in the 2016 u.s. Presidential election. *Social Psychological and Personality Science*, 10(5): 643–652.
- World Bank. 2023.
- Wüthrich and Zhu. 2023. Omitted Variable Bias of Lasso-Based Inference Methods: A Finite Sample Analysis. *The Review of Economics and Statistics*, 105(4): 982–997.
- Yourish, Buchanan, and Liu. 2021. The 147 republicans who voted to overturn election results. *The New York Times*. <https://www.nytimes.com/interactive/2021/01/07/us/elections/electoral-college-biden-objectors.html>.
- Ziblatt and Levitsky. 2018. How democracies die. *Crown*.