Author's response to reviews

Title: The influence of societal individualism on a century of tobacco use: modelling the prevalence of smoking

Authors:

John C Lang (j8lang@uwaterloo.ca)
Daniel M Abrams (dmabrams@northwestern.edu)
Hans De Sterck (hdesterck@uwaterloo.ca)

Version: 4 Date: 29 October 2015

Author's response to reviews: see over
Dear Reviewers and Editors of BMC Public Health,

Please find enclosed below our responses to the reviewers’ comments.

- We respond directly to reviewer commend in red on pages ii-v.
- We include a copy of our manuscript and appendices showing the tracked changes made on pages 1-18 (manuscript), A.1-A.5 and B.1-B.2 (appendices). New text is shown in blue and removed text is shown in red.
- References to page and line numbers in our comments on pages ii-v refer to the tracked changes version of our manuscript and appendices, i.e. to pages 1-18, A.1-A.5, and B.1-B.2 that immediately follow pages ii-v.

We have uploaded four new files to the online portal:

- A new manuscript file (lang_smoking_revised_final.tex) – without tracked changes,
- A new bbl file (lang_smoking_revised_final.bbl)
- A new Additional File 1 (Additional File 1_revised_final.tex) – without tracked changes, and
- An Additional File 5 (contains the Matlab Code used to produce this manuscript, see comment #4 of Reviewer #2).

Sincerely,

John C. Lang (corresponding author)
Department of Applied Mathematics
University of Waterloo
200 University Avenue West
Waterloo, Ontario, N2L 3G1, Canada.
Email: j8lang@uwaterloo.ca
Phone: +1-224-261-3780
Fax: +1-519-746-4319

Daniel M. Abrams
Office M444
Technological Institute
Northwestern University
2145 Sheridan Road
Evanston, IL, 60208-3125, USA
Email: dmabrams@northwestern.edu
Phone: +1-847-491-5346
Fax: +1-847-491-2178

Hans De Sterck
Department of Applied Mathematics
University of Waterloo
200 University Avenue West
Waterloo, Ontario, N2L 3G1, Canada.
Email: hdesterck@uwaterloo.ca
Phone: +1-519-888-4567 ext. 37550
Fax: +1-519-746-4319
Reviewer #1:

This is a well-written, novel contribution to the tobacco use modelling literature. Whereas previous models of tobacco use have relied on micro-level data on smoking behavior, the authors use a theory-driven approach to consider the broader social mechanism by which individuals choose to initiate or quit smoking in modeling the smoking epidemic across different countries. It is impressive that their overall predictions about smoking prevalence in relation to collectivism/individualism corresponds well with the data. The authors adequately address important limitations of their model, including its inability to account for population heterogeneity and the role of social networks in the utility function.

Thank you for the positive comments.

Major Compulsory Revisions

1. The diagram in Figure 2 is rather convoluted and difficult to follow with the text. For example it is unclear what is meant by some of the lines and arrows displayed until much later in the paper. The initial description for Figure 2 (p. 3 line 29 – p. 4 line 2) lists steps that do not explicitly correspond with what is displayed in the figure. (Why is “Tobacco use data” linked to “IDV”? Does “Tobacco use data” predict the relationships in phases ii and iii? Or does the box for ‘predict’ only refer to the smoking prevalence model?)

We agree it is difficult to summarize our theory-driven methodology early on in the paper, but we find it important to do so (because our methodology is different from what exists in the literature). We agree that Fig. 2 requires more context when it is discussed in the introductory section. In the revised manuscript we have done our best to add the missing context by clarifying Fig. 2 in an expanded caption (lines 1-15 on page 17).

The paper would be improved by moving some of the writing describing the model phases from the Results section (p.8-9) into the Methods section.

Thank you for the helpful suggestion. We have reorganized the Results and Methods sections. A description of our procedure for testing model predictions is given in the Methods section from line 11 (page 8) until line 9 (page 10). A summary of the results is given in the Results section from line 11 (page 10) until line 3 (page 12).

2. As an alternative to the diagram, Figure 2 could be replaced with a table outlining specific steps in the modeling process for each phase: Model specification, Phase (i), Phase (ii), and Phase (iii). This would clarify the stages of model development and improve readability, especially for readers of the journal who lack modeling backgrounds.

Thank you for the suggestion for an alternative to Fig. 2. It is the authors’ preference to keep Fig. 2, which we believe is much improved by the previous comment.

Minor Essential Revisions

3. Figure 2 is first introduced with no further explanation (p. 2 line 33) until later (p. 3 line 29 – p. 4 line 2). Consider moving “Figure 2 summarizes our approach graphically” on page 1 line 33 to page 3 line 29.

We agree that line 32 (page 3) is a more appropriate place to first reference Fig. 2 and have made the appropriate changes.

Discretionary Revisions

4. The authors use publications on the health effects of smoking as a proxy for utility derived from smoking (page 4 line 45). The authors could also describe how the decline in individual utility over time (discounting as shown in Figure 3b) corresponds with the general trend in growth in the number of
scientific articles being published overall and with ‘tobacco fatigue’ in recent decades. That is, the tendency for the public to derive increasingly less utility from new information about smoking over time due to knowledge saturation.

We have expanded our discussion of the relationship between individual utility, public knowledge about the health effects of smoking, and number of publications on lines 20-28 (page 5).

5. In Appendix B, additional information about sensitivity analyses conducted would improve model transparency. The authors state that several model assumptions were tested—which assumptions in particular? What ‘various’ combinations of local and global parameters were investigated, and what were the specified bounds?

We tested assumptions about the functional form of the utility function (step utility function versus the utility function given in Eq. 2) and which variables were local and which were global. Simulations were performed for each combination of utility function and local/global variable specification. We have given additional details on page B.2.
Reviewer #2

The study is on an interesting topic and it is fairly conceived with basic novelty to be published on BMC Public Health. However, I have several concerns about the manuscript which are outlined below.

We thank you for the positive comments.

1. The theoretical framework could be more thoroughly developed and the methodology, especially the social utilities scheme should be explained in a clear way, and for a non-specialist audience. While the method is not completely new and a list of other publications are cited to follow up for getting clear understanding about the model, nevertheless needs to have enough exposition of the method so that the interested general reader can get some better understanding. The current version of this paper somewhat fails in this regard. Just for an example, after immediately the Eq (1), the representation of a and t should be mentioned with its values.

We have added the explicit definition of the variable t on line 29 (page 4). We have reorganized our discussion of the parameter a so that it comes immediately after the introduction of Eq. (1) from line 31 (page 4) to line 6 (page 5). Our discussion has also been expanded to include a brief discussion of the expected range of parameter a on lines 2-6 (page 5).

2. The assumption of an individual's decision to smoke is based on the desire to maximize total utility (that includes individual utility and social utility) seems okay, but the concept of measuring the social utility is not very openly illustrated. Typically, the social utility perhaps link with complex social interactions of smokers.

We agree that social utility follows from complex social interactions (line 8-11, page 13), and we model it (in the case of total utility from smoking) as an increasing function of x, i.e. the fraction of the population that smokes. This is consistent with a “mass action” law approach and appears to work for the dynamics of human social groups in several contexts. We have tried to clarify this earlier on in the text on lines 9-13 (page 5).

3. There needs to be a simple and clear explanation of the Hofstede's IDV in the context of this research. How the Hofstede's IDV measure has incorporated here?

The relative conformity parameter a is interpreted as the parameter in our model that reflects a society’s level of individualism/collectivism. To test this prediction we compare the values of a (obtained from fitting the model to smoking prevalence estimates) to an established measure of societal individualism, i.e. Hofstede’s IDV (Phase (ii)). The relative conformity parameter a also plays a central role in determining the average slope s_x and the peak year t_max, therefore, we also compare the values of s_x and t_max with IDV (Phase (iii)). We attempt to clarify these details by expanding the caption of Fig. 2 on lines 1-15 (page 17), by providing a detailed explanation of Phases (ii)-(iii) in the Methods section from line 13 (page 9) to line 9 (page 10), and by rewording our introduction from line 37 (page 3) to line 13 (page 4).

4. Although the model fitting section is depicted well, all simulation code and necessary data sets could be available (as an archive) to readers for use.

We are happy to include the simulation code as Additional File 5, see lines 10-11 (page 9) and lines 29-30 (page 18).

Moreover, in page 7, line 41, the equation should be corrected to Ei,2 = ||X #i2 – X #i2||^2 and then in line 43, E2 = #_i^?Ei2. similar should be applied to line 45.

Thank you for pointing out that this notation is unclear. We attempt to clarify our notation on lines 29-31 (page 8). We have corrected the typo on lines 32-33 (page 8), thank you for taking the time to carefully read our manuscript.
How the “length of X #i” has been estimated is not clear at all!

Lines 18-19 (page 8) define $x^i(t)$ as the smoking prevalence estimated for country $i$ at time $t$. Lines 19-20 (page 8) defines $X^i$ as a vector with elements $x^i(t)$ for all times $t$ where smoking prevalence estimates $x^i(t)$ are available. The length of $X^i$ is equal to the number of years for which smoking prevalence estimates are available for country $i$, i.e. the length of $X^i$ is calculated directly and does not have to be estimated. We clarify this on lines 7-10 (page 9).

5. There needs to be a much better explanation of the results which are very abridge in the current version. Perhaps needs some more highlights on the significant findings.

We have attempted to improve the explanation of our results by expanding the caption of Fig. 2 on lines 1-15 (page 17), adding a section to Methods (Testing the Model) that explains Phases (i)-(iii) in detail from line 11 (page 8) to line 9 (page 10), and by adding a prefatory statement to the Results section explaining how the results we present provide evidence in support of the model we developed, see lines 12-22 (page 10). We have also included an additional line in the conclusion on lines 10-13 (page 14).

Also, in page 9, line 40, it would be “... from 1966 to 1973 (see Fig. ...)

The full statement is “the measurement for peak year in the UK would slightly decrease from 1973 to ~1966”. We have clarified our notation so that the statement now reads “the measurement for peak year in the UK would slightly decrease from $t_{max} = 1973$ to $t_{max} \approx 1966$”, see lines 43-46 (page 11).

6. Discussion section is nicely written. All limitations seem realistic and some of them are easily justifiable. However, the alternative explanations in page 10 make things somewhat ambiguous.

Thank you for the positive comment. We think that the alternative explanations are mostly accounted for implicitly in our modelling framework, and hence, the effect of confounding variables on our results is limited. We have attempted to clarify this argument and make it more explicit on lines 15-18 (page 12) and lines 37-40 (page 11).

7. Finally, the authors may test their model with data from some non OECD or developing countries. Perhaps this could be a future research agenda.

Thank you for your suggestion. We have incorporated this suggestion on line 8 (page 13).
The influence of societal individualism on a century of tobacco use: modelling the prevalence of smoking

John C Lang1*, Daniel M Abrams2† and Hans De Sterck1‡

Abstract

Background: Smoking of tobacco is estimated to have caused approximately six million deaths worldwide in 2014. Responding effectively to this epidemic requires a thorough understanding of how smoking behaviour is transmitted and modified.

Methods: We present a new mathematical model of the social dynamics that cause cigarette smoking to spread in a population, incorporating aspects of individual and social utility. Model predictions are tested against two independent data sets spanning 25 countries: a newly compiled century-long composite data set on smoking prevalence, and Hofstede’s individualism/collectivism measure (IDV).

Results: The general model prediction that more individualistic societies will show faster adoption and cessation of smoking is supported by the full 25 country smoking prevalence data set. Calibration of the model to the available smoking prevalence data is possible in a subset of 7 countries. Consistency of fitted model parameters with an additional, independent, data set further supports our model: the fitted value of the country-specific model parameter that determines the relative importance of social and individual factors in the decision of whether or not to smoke, is found to be significantly correlated with Hofstede’s IDV for the 25 countries in our data set.

Conclusions: Our model in conjunction with extensive data on smoking prevalence provides evidence for the hypothesis that individualism/collectivism may have an important influence on the dynamics of smoking prevalence at the aggregate, population level. Significant implications for public health interventions are discussed.

Abbreviations: Organization for Economic Co-operation and Development (OECD), Hofstede’s Individualism Index (IDV)

Keywords: smoking; individualism; mathematical modelling; social dynamics; non-infectious diseases

Background

In the fifty years since the first report of the Surgeon General’s Advisory Committee on Smoking and Health [1] the smoking epidemic has been responsible for more than 20 million deaths in the United States alone [2, 3], and continues to be responsible for over 6 million deaths worldwide each year [4, 5]. The strong social component of the dynamics of smoking prevalence has been modelled mathematically [6–10] and examined statistically through analysis of social network data [11] and survey data [12–14]. However, whereas previous works tend to focus on the micro-level, in
In this paper we investigate how social aspects of smoking affect its prevalence at the societal level.

Significant inter-country differences exist in smoking prevalence [15]. For example, Fig. 1 shows smoking prevalence estimates over most of the past century for Sweden and the USA, obtained from smoking prevalence surveys and cigarette consumption data (collectively referred to as tobacco use data, see Data subsection in Methods). In both countries, smoking prevalence increased rapidly starting from the early decades of the 20th century and reached a peak in the 1960s–1980s era when the adverse health effects of smoking became widely known [1], after which smoking prevalence declined rapidly. However, there are conspicuous differences between the curves: the rate of smoking adoption and cessation before and after the peak is much greater in the US than in Sweden, and the peak in prevalence in the US occurs much earlier than in Sweden.

Considerable time and resources have been devoted to identifying the factors that contribute to smoking prevalence. Major factors include differences in beliefs about the harm of smoking [16], socio-economic status [17, 18], cost [19], regulation/tobacco control policies [20–22], and gender [23]. However, we note that these advances in the understanding of the factors contributing to smoking prevalence are based primarily on micro-level data, methods that inform general hypotheses, and non-mathematical descriptive models. Indeed, comprehensive and quantitative cross-national analyses of how all these factors affect smoking prevalence are rare [15]. Existing studies that compare national trends in smoking prevalence, as well as the factors that contribute to these trends, tend to take a descriptive [24, 25] and/or statistical [15] approach, and do not address the mechanism underlying the key decision of whether or not to smoke in a quantitative manner [26].

In this paper we present a new model for the social spreading of smoking. We aim to create and test a tractable mathematical model, that is, a model for qualitative dynamics from which insight (including causation) can be drawn. This differs from the statistically-driven approach often used in areas such as econometrics and medicine, where correlations may be uncovered and analyzed without formulating first-principle-based dynamic mathematical models. The statistical approach is difficult to apply here because the amount of available data on historical smoking dynamics is small. Our model-based approach has much in common with simple explanatory mathematical models that have been successful in, e.g., epidemiology and population dynamics. Figure 2 summarizes our approach graphically.

Our model incorporates the concepts of individual utility from smoking, i.e. the utility an individual derives directly from the act of smoking (including awareness of health effects), and social utility from smoking, i.e. the utility an individual derives indirectly from smoking through social interactions with other smokers (peer influence and social inertia). Together these two quantities determine the total utility from smoking. Our model assumes that an individual’s decision to smoke is based on the desire to maximize total utility. By invoking this decision-making mechanism in a simple mathematical model, our approach differs from the approaches of the previous mathematical [7–10] and descriptive/statistical [15, 24–26] models. Whereas previous mathematical models generally require the calibration of many parameters (leading to difficulties in analysis, interpretation, and overfitting), we
propose a simple approach based on principles of social psychology and sociol-
ogy whose predictions can be directly compared to smoking prevalence tobacco use data. Whereas previous descriptive and statistical models lack an underlying decision-making mechanism, we propose a model with a decision-making mechanism that is capable of incorporating factors previously identified as contributing to smoking prevalence. Specifically, we note that monetary cost, beliefs about the harm/health effects of smoking, and regulation/tobacco control policies are all implicitly accounted for in the concept of individual utility from smoking. Our simple model applies to the population level, focusing on major effects that may influence the temporal dynamics of smoking across societies. It proposes a mechanism for smoking adoption and cessation that hinges on the balance between individual and social utility (which both encompass other more fine-grained factors). Matching the model to real-world data reveals that the balance between social and individual utility indeed is an important factor in the temporal dynamics of smoking, differentiating between countries in a way that is consistent with known measures of societal individualism. This lends support to the compelling hypothesis that the balance between individual and social utility, which we will show to be related to societal individualism, is indeed an important society-level driver for the temporal dynamics of smoking prevalence. This is consistent with previous findings that the level of individualism/collectivism of a society may have fundamental implications for its biology [27, 28], as well as its behaviour [12, 29–32].

The model we propose is explained in the Model Specification subsection of the Methods section below. In the context of societal individualism/collectivism, the parameter in our model that controls the relative importance of individual versus social utility is interpreted as follows: the greater the relative contribution of individual utility to total utility (at the expense of social utility), the more individualistic the society is interpreted to be. Conversely, the greater the relative contribution of social utility to total utility (at the expense of individual utility), the more collectivistic the society is interpreted to be. As described in detail in the Results section below (see Testing the Model subsection in Methods), this allows us to test the model’s predictions against independently collected smoking prevalence tobacco use and individualism/collectivism data sets in three separate phases, see Fig. 2. First, using tobacco use data we compile smoking prevalence data estimates spanning the past century for seven countries belonging to the Organisation for Economic Co-operation and Development (OECD) and find good agreement between these estimates and the fitted model (Phase (i) in Model + Testing, see Fig. 2). Second, the country-specific parameter in our model that controls the relative importance of individual versus social utility, i.e. the parameter that we interpret as the degree of societal individualism/collectivism (the relative conformity parameter \(a\), see Model Specification subsection in Methods section), and that we fit to smoking prevalence data estimates, is found to be significantly correlated to an established measure of societal individualism for each country (Hofstede’s IDV [33]), in agreement with the predictions of the model (Phase (ii) in Model Testing, see Fig. 2). Thirdly, given the predicted relationship between the relative conformity parameter \(a\) and Hofstede’s IDV (tested in Phase (ii)), and given the central role played by societal individualism/collectivism in the relative conformity parameter \(a\)
in our model, we are motivated to investigate directly the role that individualism (as measured by Hofstede’s IDV) plays in observed historical tobacco use data. Specifically, our model predicts that more individualistic societies will show faster adoption and cessation of smoking. We investigate this in historical tobacco use data, and find that IDV is significantly correlated to the average rate of increase in smoking prevalence \( s_x \) in seven OECD countries for which historical smoking prevalence estimates are available, and that it is significantly correlated to the peak year of tobacco consumption \( t_{max} \) for 25 countries in which tobacco consumption data are available, in agreement with model predictions (Phase (iii) in Fig. 2). These findings are interpreted according to our modelling framework, and provide evidence for the compelling hypothesis that individualism/collectivism has an important influence on the dynamics of smoking prevalence at the aggregate population level.

15 Methods

Model Specification

We begin formulating our model by observing that individuals derive utility from smoking via two mechanisms. First, they derive utility directly from the act of smoking (individual utility). Second, they derive utility from social interaction with other smokers (social utility). We note that social utility commonly manifests itself in the form of peer influence or peer pressure [34, 35]. We then proceed using a modelling framework that explicitly accounts for the effect of competition between individual and social utilities, and that was first applied to explore the temporal dynamics of language death and religious affiliation as binary choice problems [36, 37]. Specifically, we propose the model

\[
\frac{dx}{dt} = b \left[ (1 - x) x^a u_x - x (1 - x)^a (1 - u_x) \right],
\]

where \( x = x(t) \in [0, 1] \) is the fraction of smokers in the population (i.e., the prevalence) at time \( t \), \( u_x \in [0, 1] \) is the individual utility from smoking, and the constant \( b > 0 \) determines the timescale of the equation. The interpretation of the positive term in Eq. (1), which models smoking adoption, is therefore that non-smokers take up smoking at a rate proportional to the total utility derived from smoking, \( x^a u_x \), which is the weighted product of the individual utility from smoking \( u_x \) and the social utility from interactions with other smokers \( x \), with weighting determined by the constant parameter \( a \). Since societies with large \( a \) weigh changes in social utility more heavily than changes in individual utility when calculating total utility, we call \( a \) the relative conformity parameter. We therefore interpret societies with large \( a \) to be more collectivistic (or less individualistic) than societies with small \( a \). Conversely, the interpretation of the negative term in Eq. (1), which models smoking cessation, follows analogously: smokers cease smoking at a rate proportional to the total utility derived from non-smoking, \( (1 - x)^a (1 - u_x) \), which is the weighted product of the individual utility from non-smoking \( u_y = 1 - u_x \) and the social utility from interactions with other non-smokers \( 1 - x \), where we have normalized individual utilities from smoking \( u_x \) and from non-smoking \( u_y \) such that \( u_x + u_y = 1 \). Since societies with large \( a \) weigh changes in social utility more heavily
than changes in individual utility when calculating total utility, we call $a$ the relative conformity parameter. We therefore interpret societies with large $a$ to be more collectivistic (or less individualistic) than societies with small $a$. In other words, since total utility, we expect strongly collectivistic societies to have $a > 1$ and strongly individualistic societies to have $a < 1$. We note that this modelling framework is conceptually consistent with the findings presented in [12]: that personal attitudes about smoking have a stronger influence on smoking behaviour in individualistic countries than in collectivistic countries. We also note that, although social utility follows from complex social interactions, we have made the simplifying assumption that the social utility of a group, e.g. of smokers, is proportional to the size of that group, e.g. the smoking prevalence $x$. This assumption has been shown to work well in previous works [36, 37].

Next, we observe that a combination of factors, including advances in our understanding of the health effects of smoking and public policy initiatives designed to curb smoking, have likely reduced individual utility from smoking ($u_x$) over the past century. Thus, in a significant departure from previous work that treats individual utility as a constant [36, 37], we account for this decline in individual utility by using the cumulative number of scholarly articles on the health effects of smoking ($n(t)$) as a proxy for the reduction in individual utility over the past century. Since each additional article represents an increase in the public knowledge about the health effects of smoking, we assume that individual utility decreases with each additional article published. We also assume that public knowledge about the health effects of smoking becomes saturated after a large number of articles have been published. In other words, we assume that public knowledge about the health effects of smoking is subject to diminishing marginal returns from additional articles published, and hence, individual utility is subject to diminishing marginal losses from additional articles published. Specifically, we apply these assumptions by following the principle of temporal discounting [38], i.e. we assume that each additional article published is discounted by the factor $\delta \in [0, 1]$ so that for year $t$

$$u_x(t) = u_\infty + \delta^{n(t)}(u_0 - u_\infty),$$

where $u_0$ and $u_\infty$ are the limiting individual utilities from smoking when there is no knowledge and perfect knowledge of the adverse effects of smoking, respectively. Here, $u_0$, $u_\infty$ and $\delta$ are parameters to be fitted to observational data.

We remark that this approach leads to better fits between model output and observational data than alternatives that do not directly take into account the effect of increased scientific understanding of health effects. For example, using the discounting formula of Eq. (2) produces a better fit (significantly lower total error $E_2$) than either constant utility $u_x(t) \equiv u_x$ or step-function utility

$$u_x(t) = \begin{cases} 
    u_0 & \text{if } t < t^* \\
    u_\infty & \text{if } t \geq t^*
\end{cases},$$

where $t^*$ is a threshold parameter whose value is determined by the fitting procedure. Note that when $u_0 > u_\infty$ the step-function utility is consistent with the expectation
that increasing knowledge of health effects has indeed influenced the individual utility from smoking over the past century.

4 Data

We note that Eq. (1) subject to Eq. (2) requires the fitting of four parameters per country \( x_0 = x(t_0), a, u_0, \) and \( u_\infty \) for all countries in the data set (see Model Fitting subsection in Methods section). We determine these parameters by fitting them to estimated historical smoking prevalence data and proxy data on the health effects of smoking. We summarize the methods used to obtain these data below. Note: No human subjects participated in this study. No consent was necessary to obtain.

Tobacco Use Data: Smoking Prevalence and Cigarette Consumption Data

We consider smoking prevalence \( x(t) \in [0, 1] \) for 24 OECD countries which we download from the OECD iLibrary online statistical database [39] in Excel format. We also consider manufactured cigarette consumption (in grams) per person per day \( c(t) \) for the same 24 OECD countries plus Romania (which is a non-OECD country) [40, 41]. When available, cigarette consumption data is downloaded directly from the International Smoking Statistics (Web Edition) website [40] in Excel format. Cigarette consumption data for countries not included in the International Smoking Statistics (Web Edition) are retrieved from the International Smoking Statistics (2nd Ed.) [41] by manually transferring these entries into Excel. We refer to smoking prevalence and cigarette consumption data collectively as tobacco use data. We make these data available in CSV format in an additional file (see Additional File 2.csv), which contains four columns: country number as it appears in Table A.1 of the additional file Additional File 1.pdf, year \( t \), measurement \( x(t) \) or \( c(t) \), and type of measurement (0 indicates a smoking prevalence measurement, while 1 indicates a cigarette consumption measurement).

Unfortunately smoking prevalence data is limited to, on average, only 21.5 observations over a period of 31.4 years spanning 1960–2012 [39]. As such, it misses much of the crucial period in the earlier parts of the 20th century during which smoking steadily gained popularity in many countries. However, historical national cigarette consumption data is available for the same 24 OECD countries plus Romania for an average of 78.4 observations over a period of 82.2 years spanning 1900–2012 [40, 41]. Since our model is specified in terms of smoking prevalence, we estimate smoking prevalence from cigarette consumption in order to exploit the much richer cigarette consumption data for model fitting purposes. First, we assume a linear relationship between smoking prevalence \( x(t) \) and smoking consumption \( c(t) \):

\[
 x(t) = Cc(t) + B. \tag{3}
\]

Next, we calculate estimates \( \hat{C} \) and \( \hat{B} \) by regressing smoking prevalence \( x(t) \) on tobacco consumption \( c(t) \) for all years for which both measurements are available. The results of this regression are summarized in Table A.2 of the additional file Additional File 1.pdf, which illustrates that the assumption that \( x \) and \( c \) are linearly related does not hold equally well for all countries. In order to restrict ourselves to
the cases where the assumption of linearity between \(x\) and \(c\) is valid we restrict\(^1\) ourselves to the seven OECD countries with \(R^2 \geq 0.7, p < 0.001\), and \(n_{\text{obs}} \geq 15\):\(^2\) Australia, Canada, France, New Zealand, Sweden, the United Kingdom, and the\(^3\) United States. We display the raw data for these seven OECD nations in Fig. A.1\(^4\) of the additional file Additional File 1.pdf. The smoking prevalence for these seven\(^5\) OECD countries is then estimated from tobacco consumption using the relationship\(^6\)
\[
\hat{x}(t) = \hat{C}c(t) + \hat{B}.
\] (4)

Note that survey-based prevalence data are susceptible to noise stemming from\(^11\) variations in the survey methodology. In particular, prior to performing the linear\(^12\) regression of \(x\) on \(c\) for France, we removed the outlier \(x(1960) = 0.32\) since it is\(^13\) inconsistent with the rest of the data for France, see Fig. A.1(c) of the additional\(^14\) file Additional File 1.pdf. Specifically, the Grubbs test on \(x/\hat{x}\) indicates that the\(^15\) 1960 data point is a significant outlier \((p < 0.05)\). This can also be seen intuitively:\(^16\) from \(t = 1960\) until the next measurement at \(t = 1965\) smoking prevalence drops\(^17\) from \(x(1960) = 0.32\) to \(x(1965) = 0.25\) (a decrease of 21.9\%), while cigarette\(^18\) consumption steadily increases from \(c(1960) = 3.6\) to \(c(1965) = 4.1\) (an increase\(^19\) of 13.9\%). Given the population in France in 1960 (45.5 million) and in 1965 (48.6\(^20\) million) [42], this would correspond to an increase in the average mass of cigarettes\(^21\) smoked (in grams) per smoker per day from 11.3 to 16.4 (an increase of 45.1\%) over\(^22\) a short 5 year period. This is in sharp contrast with the relatively stable relationship\(^23\) between \(x\) and \(c\) for France’s remaining data points and justifies the exclusion of the\(^24\) outlier \(x(1960) = 0.32\). With the outlier removed, France satisfies our data quality\(^25\) requirements for inclusion in the set of seven OECD countries \((R^2 \geq 0.7, p < 0.001)\).\(^26\) Our assumption of linearity between smoking prevalence \(x\) and cigarette consumption \(c\) is not perfect, but it appears to be satisfied at most times in countries\(^28\) where both data sets are available. Quadratic or other higher order terms could be\(^30\) included, but additional unknown parameters would have to be introduced and the\(^31\) limitations of our data set (sparsity, noise) mean that there would be little or no\(^32\) improvement in the model’s fit.

Proxy Data \(n(t)\): articles published on the health effects of smoking

We calculate the cumulative number of articles published on the health effects of\(^36\) smoking \(n(t)\) by performing a search of the online research database Scopus for\(^37\) papers with

(i) tobacco, smok*, or cigar* in the title, and

(ii) death, illness, mortality, risk*, tumour*, tumor*, or cancer in the title, and

(iii) medicine, dentistry, nursing, veterinary, health professions, or multidisciplinary in the subject area, and

(iv) plant*, mosaic, botany, smog, fog, and soot not in the title.

Items (i)-(iii) are search terms included in order to select for papers researching the health effects of smoking, whereas items (iv) are search terms excluded in order to\(^45\) prevent selection of papers researching the tobacco mosaic virus (plant*, mosaic,
Testing the Model

Fitting

We fit Eq. (1) to the estimated prevalence, \( \hat{x}(t) \). To reduce the dimensionality of the optimization problem, we assume that certain universal parameters are constant across countries. Specifically, we assume that \( b \) and \( \delta \) are universal parameters, and that \( x_i(t_{i,0}) = x_{i,0}, a_i, u_{i,0}, \) and \( u_{i,\infty} \) are local parameters for country \( i \), where \( t_{i,0} \) is the first year for which cigarette consumption data \( c \) is available. We denote the smoking prevalence estimated above for country \( i \) at time \( t \) by \( \hat{x}_i(t) \). The time series of estimated smoking prevalences for country \( i \) is then denoted by the vector \( \hat{X}_i \). Analogously, we denote the time series of smoking prevalences predicted by Eq. (1) for country \( i \) by \( \tilde{X}_i \). We solve Eq. (1) using the Matlab differential equation solver ode45.

Using the Matlab function \textit{lsqcurvefit} we proceed as follows:

1. Holding universal parameters constant, for each country \( i \) we find the \( x_{i,0}, a_i, u_{i,0} \), and \( u_{i,\infty} \) that minimize

\[
E_{i,2} = \| \tilde{X}_i - \hat{X}_i \|_2^2,
\]

where the \( L_2 \) norm \( \| \cdot \|_2 \) for a vector \( y = (y_1, \ldots, y_n) \in \mathbb{R}^n \) is given by

\[
\|y\|_2 = \sqrt{\sum_{j=1}^{n} y_j^2}.
\]

2. Holding local parameters constant for each country \( i \), we find the \( b \) and \( \delta \) that minimize \( E_2 = \sum_i \| \tilde{X}_i - \hat{X}_i \|_2^2 = \sum_i E_{i,2} \).

3. Repeat steps (1) and (2) until either

(a) the change in the objective function \( E_2 = \sum_i E_{i,2} \) is below tolerance \( tol \),

or

(b) the number of iterations exceeds a limit \( max_{itn} \).

We perform the optimization with the initial guess \( u_{i,0} = 0.51, u_{i,\infty} = 0.49, a_i = 1, b = 1, \) and \( \delta = 0.9985 \). We also provide the optimization algorithm \textit{lsqcurvefit} with constraints

\[
0 \leq a_i, b \leq 2 \text{ and } 0 \leq x_{i,0}, u_{i,0}, u_{i,\infty}, \delta \leq 1,
\]

and with parameters \( tol = 10^{-6} \) and \( max_{itn} = 150 \). The fitting procedure terminates after 114 iterations, the results of which are recorded in Table 1 and Fig. A.2.
of the additional file Additional File 1.pdf. For completeness, Table 1 also records\(^1\) the average of the absolute value of the difference between \(\hat{X}_i\) and \(\tilde{X}_i\):

\[
E_{i,1} = \frac{\|\tilde{X}_i - \hat{X}_i\|_1}{\text{length of } \hat{X}_i},
\]

where the \(L_1\) norm \(\|\cdot\|_1\) for a vector \(y = (y_1, \ldots, y_n) \in \mathbb{R}^n\) is given by \(\|y\|_1 = \sum_{j=1}^{n} |y_j|\), and where the length of \(\hat{X}_i\) is equal to the number of elements of \(\hat{X}_i\), i.e. the length of \(\hat{X}_i\) is equal to the number of years for which smoking prevalence estimates \(\hat{x}_i(t)\) are available. For \(\mathcal{C}\) complete model simulation code with all necessary data files are available upon request, see Additional File 5.zip.

Phase (ii): Test of model implications for \(a\)

If the model and its interpretation are correct and the balance between individual and social utility is a relevant factor for the temporal dynamics of smoking prevalence, then we expect that the fitted relative conformity parameter \(a\) will be different for different countries and will capture something meaningful about the individualism/collectivism of a society. To test this we compare with Hofstede’s IDV, an established metric for societal individualism [33] that has been evaluated in most countries. Specifically, by computing the linear regression of \(a\) on IDV we expect to reveal a significant negative correlation between these two quantities (negative because \(a\) increases with collectivism while IDV decreases with it).

Phase (iii): Test of model implications for slope and peak year

Besides the correlation of \(a\) with collectivism, we note that another prediction is implicit in model (1). As the relative conformity parameter \(a\) increases, the model requires that changes in smoking prevalence occur more slowly (this is true for solutions to Eq. (1) for the range of \(a\) and \(u\) values corresponding to the observational data). Put another way, societies with higher levels of individualism should experience faster changes in smoking prevalence. Intuitively, when smoking prevalence is low the lack of existing smokers inhibits smoking initiation more strongly in a collectivistic society than in an individualistic society. Thus, we expect the average rate of increase in a collectivistic society to be smaller than in an individualistic society. In contrast, when smoking prevalence is high, and once the deleterious health effects of smoking become widely known and negatively impact individual utility from smoking, the presence of existing smokers inhibits smoking cessation more strongly in a collectivistic society than in an individualistic society. In both cases collectivism acts as a brake on change in the status quo (higher cultural inertia [43, 44]). Specifically, we expect the average slope \(s_x\) of the smoking prevalence curves leading up to the peak smoking prevalence increases with Hofstede’s IDV and decreases with \(a\), respectively. Here we define the average slope \(s_x\) to be

\[
s_x = \frac{\hat{x}(t_{max}) - \hat{x}(t_0)}{t_{max} - t_0},
\]

where \(t_0 = 1920\) is the first year for which smoking prevalence estimates are available in the subset of seven OECD countries, and where \(t_{max}\) is the earliest
year for which the maximum tobacco consumption was recorded, see Table A.3 of
the additional file Additional File 1.pdf.

This reasoning further suggests that the peak year for smoking prevalence $t_{max}$
should be later in collectivistic societies and earlier in individualistic societies.
Specifically, we expect $t_{max}$ to be significantly negatively correlated with IDV and
significantly positively correlated with $a$. Note that our assumption of a linear rel-
ationship between national cigarette consumption and smoking prevalence is not
needed to establish $t_{max}$, so the relationship between $t_{max}$ and IDV is independent
of any model assumptions.

Results: Testing the Model

We now test the model in three phases, as depicted in Fig. 2. In Phase (i) we cali-
brate the model using smoking prevalence estimates $\hat{x}_i(t)$ derived from the tobacco
use data (smoking prevalence and cigarette consumption data). The model predic-
tions about how the relative conformity parameter $a$, the slope $s_x$ and the peak year
$t_{max}$ are related to the level of individualism/collectivism in society are tested in
Phases (ii)-(iii) by comparison to an existing measure of individualism/collectivism,
I.e. to Hofstede’s IDV. Since the model is calibrated using one set of data (smoking
prevalence and cigarette consumption data) and its predictions are verified using a
separate data set (Hofstede’s IDV), Phases (i)-(iii) provide significant evidence in
support of the model that we developed in Eq. (1).

Phase (i): Direct test

Figure 1 shows the fit of our model to data sets from the United States and Sweden
(additional fits and parameter values are displayed for our set of seven OECD
countries in Fig. A.2 of the additional file Additional File 1.pdf and Table 1). The
average of the absolute value of the difference between smoking prevalence estimates
$\hat{x}$ and the output of Eq. (1) ranges from a low of 0.005 for France to a high of 0.018
for the United Kingdom (see $E_{i1}$ in Table 1). The good agreement that we found
with all data sets provides support for the model.

Phase (ii): Test of model implications for $a$

If the model and its interpretation are correct and the balance between individual
and social utility is a relevant factor for the temporal dynamics of smoking
prevalence, then we expect that the fitted “relative conformity parameter” $a$ will
differ for different countries and will capture something meaningful about
the individualism/collectivism of a society. To test this we compare with Hofstede’s,
IDV, an established metric for societal individualism [33] that has been evaluated in
most countries. Panel (a) of Fig. 4 shows the comparison between the fitted $a$
values and IDV. As expected, the relative conformity parameter $a$ shows significant
differences for different countries, and, as expected, a and IDV are significantly negatively
correlated with Hofstede’s IDV (negative because $a$ increases with collectivism while
IDV decreases with it) (see Table 2). This concordance with independently assessed
individualism values supports our model.
Phase (iii): Test of model implications for slope and peak year

Besides the correlation of α with collectivism, we note that another prediction is implicit in model (1). As the relative conformity parameter increases, the model requires that changes in smoking prevalence occur more slowly (this is true for solutions to Eq. (1) for the range of α and α values corresponding to the observational data). Put another way, societies with higher levels of individualism should experience faster changes in smoking prevalence. Intuitively, when smoking prevalence is low the lack of existing smokers inhibits smoking initiation more strongly in a collectivistic society than in an individualistic society. Thus, we expect the average rate of increase in a collectivistic society to be smaller than in an individualistic society. In contrast, when smoking prevalence is high, and once the deleterious health effects of smoking become widely known and negatively impact individual utility from smoking, the presence of existing smokers inhibits smoking cessation more strongly in a collectivistic society than in an individualistic society.

In both cases collectivism acts as a break on change in the status quo (higher cultural inertia). Panel (b) of Fig. 4 and panel (a) of Fig. 5 demonstrate illustrate the relationship between $s_x$ and IDV and between $s_x$ and α: that this is indeed the case: the average slope $s_x$ of the smoking prevalence curves leading up to the peak increases with Hofstede’s IDV and decreases with α, respectively. Here we define the average slope $s_x$ to be

$$s_x = \frac{\hat{x}(t_{max}) - \hat{x}(t_0)}{t_{max} - t_0},$$

where $t_0 = 1920$ is the first year for which smoking prevalence estimates are available in the subset of seven OECD countries, and where $t_{max}$ is the earliest year for which the maximum tobacco consumption was recorded, see Table A.3 of the additional file. Figure 6 illustrates the relationship between $t_{max}$ and IDV: $t_{max}$ is significantly negatively correlated with IDV (shown) and significantly positively correlated with α (see Fig. 5(b)). Note that our assumption of a linear relationship between national cigarette consumption and smoking prevalence is not needed to establish $t_{max}$, so Fig. 6 is independent of any model assumptions. All correlations are significant (see Table 2).

We note that fluctuations in the data due to either volatility in tobacco consump-

| tion or measurement error may affect reported $t_{max}$. Smoothing of the data could be applied prior to calculation of peak year, however, the choice of smoothing algorithm is itself arbitrary and unnecessarily complicates our findings without significantly altering the result. For example, consider the seven OECD countries for which we have estimated historical smoking prevalence data $\hat{X}_i$. We observe that the model fitting procedure described in the Methods section results in the timeseries $\hat{X}_i$, which we can consider as one possible smoothing of the data $\hat{X}_i$. In this case, the measurement for peak year does not change substantially after smoothing for most countries (see Fig. A.2(a)-(e) of the additional file Additional File 1.pdf), while the measurement for peak year in the USA would slightly increase from $t_{max} = 1963$ to $t_{max} = 1967$ and the measurement for peak year in the UK would slightly decrease from $t_{max} = 1973$ to $t_{max} = 1966$ (see Fig. A.2(f)-(g) of the additional file).
These changes would result in no discernible net change in the relationship between peak year and individualism, but would result in added complexity, and hence, in a greater chance of introducing additional error. This reasoning further suggests that the peak year for smoking prevalence $t_{\text{max}}$ should be later in collectivistic societies and earlier in individualistic countries. As shown in Fig. 6, the raw observational data are consistent with this prediction: $t_{\text{max}}$ is significantly negatively correlated with IDV (shown) and significantly positively correlated with $a$ (see Fig. 5(b)). Note that our assumption of a linear relationship between national cigarette consumption and smoking prevalence is not needed to establish $t_{\text{max}}$, so Fig. 6 is independent of any model assumptions.

### Discussion and Conclusion

#### Limitations

Before discussing the limitations of our model, it is worth discussing the potential effect of confounding variables on our model. Specifically, we argue that the effect of confounding variables on our results are limited, since most potential candidates for confounding variables are actually accounted for implicitly in our modelling framework. Consider, for example, two possible candidates for confounding variables: the wealth (per capita GDP) and the strength of tobacco control policies in each country. Specifically, consider the trend in Fig. 6 where wealthier countries have, on average, earlier peak year in tobacco consumption ($t_{\text{max}}$). This relationship is easily explained by our model, since (a) our model predicts a negative correlation between peak year and Hofstede’s IDV and (b) IDV and wealth (per capita GDP) are highly positively correlated. An alternative explanation for wealthier nations having earlier peak year, however, would be that individuals who are more wealthy are better able to afford cigarettes and, in aggregate, are better able to implement strong anti-tobacco policies: in theory, the former would lead to a more rapid increase in smoking prevalence and the latter would lead to a more rapid decline in smoking prevalence. Although this alternative explanation might seem to be in competition with our model, we argue that it is in fact accounted for implicitly in our modelling framework: both wealth and the strength of tobacco control policies are contributing factors to individual utility from smoking. Furthermore, we note that although the precise timing of anti-tobacco policies is not included in the model, it is reasonable to assume that these initiatives are implemented more frequently and more intensely as the health effects of smoking are better understood - a phenomenon which is modelled using Eq. (2) and proxy data on smoking-related publications. In summary, we argue that since most potential confounding variables are actually accounted for implicitly in our modelling framework, the exposure of our results to the effects of confounding is limited.

Despite the good match between model predictions and data, a number of limitations remain. First, due to limitations in the quantity and quality of the available data, a number of limitations remain. Indeed, these changes would increase the statistical significance of our results, but again, we don’t believe that they justify the additional complexity and the introduction of additional arbitrary smoothing parameters.
smoking prevalence and tobacco consumption data, we are only able to fit the pa-
rameters in our model to seven countries, all of which have advanced/developed
economies. There is no a priori reason to believe that, given adequate sources of
data, our model would not generalize to less developed countries with lower income.

Indeed, Fig. 6 supports the position that the behaviour in less developed countries is consistent with our mathematical model. Nevertheless, our inability to directly apply our model to a larger set of more diverse countries due to a lack of good data remains a limitation of our work and an area open to future research. Sec-
ton, we have made an implicit “mean-field” approximation in taking social utility to be a function of the overall smoking prevalence $x$, rather than the local smok-
ing prevalence among contacts in an individual’s social network. Similarly, we have taken individual utility to be uniform across the population (though not in time), whereas a more detailed model might allow for variation with, e.g., gender, age and income. We note, however, that the success of our model in reproducing the trends in smoking prevalence in a manner consistent with its interpretation in the context of individualism/collectivism, despite these limitations, is generally supportive of the modelling framework we have developed. In particular, our results and the data indicate that, when averaging over gender, age and income in a country, a strong net influence remains from societal individualism on the aggregate temporal dynamics of smoking prevalence. Furthermore, if the mechanism in our model did not reflect the reality of the decision-making process for smoking then, even if it still somehow managed to fit the smoking prevalence data, we would not expect to simultaneously find correlation of (a) the relative conformity parameter $a$ with Hofstede’s individualism measure IDV (Fig. 4(a)), (b) average slope $s_x$ with IDV (Fig. 4(b)), and (c) peak year $t_{\text{max}}$ with IDV (Fig. 6). Moreover, we would not expect that the sign of these correlations would be consistent with the predictions of the model.

Our findings suggest that the correlation of individualism with faster societal change (as a consequence of a sudden change in personal utility) results from a causative influence as predicted by our model. As already mentioned, other factors such as income levels also correlate with individualism. We certainly cannot exclude that there may be other causative factors. For example, our model in its current form is incapable of explaining differences in smoking prevalence between genders and why these inter-gender differences vary between countries [15, 23]. Nevertheless, we remark that many previously proposed causative factors for differences in observed inter-country smoking dynamics can be accounted for within our modelling frame. In particular, beliefs about the harmful effects of smoking [16], the price of cigarettes [19], socioeconomic status and inequality [17, 18], and government regula-	ion [20–22] have all been cited as potential factors affecting the differences observed in inter-country smoking dynamics. Each of these factors can be interpreted within our modelling framework. For example, beliefs about the harmful effects of smoking, as well as the price of cigarettes, both likely contribute directly to individual utility derived from smoking ($u_x$) and from non-smoking ($u_y$). Moreover, socioeconomic status may affect individual utility from smoking indirectly by affecting an individual’s tolerance for risk and/or how they discount future rewards and costs (i.e. how they discount their future health status) [45]. Addressing the model’s inability to account for gender differences in smoking prevalence and explicitly quantifying
the relationship between other causative factors and model parameters in more elaborate models are potential areas for future work.

Conclusion

Despite the above mentioned limitations, the quantitative mathematical model proposed in this paper, which we derived from basic principles well-documented in the sociology and social psychology literature, appears to match real-world smoking prevalence data from a variety of countries well (to our knowledge, the largest historical data set of this type ever compiled), and all predictions of the model appear to be supported by the data. Indeed, we emphasize the strong support of the model by the data, since the model was calibrated (in Phase (i)) and its predictions were tested (in Phases (ii)-(iii)) using two separate data sets (tobacco use data and Hofstede’s IDV, respectively). In particular, the model predicts that the level of individualism or collectivism of a society may significantly affect the temporal dynamics of smoking prevalence: the strong influence of the personal utility of smoking (and its decrease due to increased awareness of adverse health effects) is predicted to lead to faster adoption and cessation of smoking in individualistic societies than in more collectivistic societies.

It has previously been argued that social support mechanisms in collectivistic societies make it more likely that a person will stop smoking [32, 46] based on findings that social support (supportive counselors) can help people to adhere to decisions to quit smoking [14]. In contrast to this behaviour at the individual level, we find that aggregate smoking prevalence decreases more slowly in collectivistic societies. Since the aggregate smoking prevalence is a function of both smoking adoption and cessation, our model suggests that this may be so because social inertia/peer pressure will either inhibit the decision to stop smoking, or encourage the decision to start smoking, more strongly in collectivistic societies than in individualistic societies.

These results suggest that it may be effective to consider culture-dependent strategies for combating the ongoing smoking epidemic. For example, interventions to discourage smoking can be tailored differently in societies or social groups whose cultures differ in how they value individualism versus collectivism [47]. Specifically, consider how the goal of many tobacco control policies is to reduce the individual utility from smoking, often by increasing the cost of cigarettes through sin taxes or by requiring warnings on cigarette packages illustrating the danger of smoking to health. Our results suggest that these tactics will be more successful in individualistic societies and less successful in collectivistic societies. In contrast, tactics that may be more successful in collectivistic societies might focus on social dangers resulting from smoking, for example by emphasizing the association between smoking and low social status [17, 18], or emphasizing the large number of individuals who have already quit. More broadly, these results demonstrate that differences in culture can measurably affect the dynamics of a social spreading process, and that a mathematical model can help to illuminate this phenomenon. We welcome future work comparing a variety of social contagion phenomena across societies. Our model suggests that the increased cultural inertia in collectivistic societies may potentially lead to slower change across a wide spectrum of spreading processes (those where important changes occur in personal utility), a hypothesis that could be supported or rejected by further study.
Competing interests
The authors declare that they have no competing interests.

Author's contributions
All authors contributed equally to this work. All authors read and approved the final manuscript.

Acknowledgements
We would like to thank Dr. James Fowler for insightful remarks on an early draft of this manuscript. DMA thanks the James S. McDonnell Foundation for support through grant #220020230. JL and HDS acknowledge support from NSERC of Canada.

Author details
1Department of Applied Mathematics, University of Waterloo, 200 University Avenue West, N2L 3G1 Waterloo, Canada. 2Department of Engineering Sciences and Applied Mathematics & Northwestern Institute on Complex Systems & Department of Physics and Astronomy, Northwestern University, 633 Clark Street, 60208 Evanston, United States.

References


http://www.census.gov/popest/data/national/totals/pre-1980/tables/popclockest.txt


http://www.census.gov/popest/data/intercensal/national/files/US-EST001NT-TOT.csv

**Figure 1** Estimated smoking prevalence in Sweden and the USA (1920-2020). Estimated smoking prevalence \( \hat{P} \) versus time for the United States (dots) and Sweden (asterisks). The solid lines give the curves of best fit for Eq. (1).
positive correlation between average slope in smoking prevalence $s$ and IDV. Linear regression of $s$ versus Hofstede’s IDV confirms that these two quantities are significantly (positively) correlated and linear regression of $t_{max}$ on IDV confirms these two quantities are significantly (negatively) correlated, see Fig 4(b) and Fig. 6, respectively. Results from Phases (i)-(iii) provide evidence in support of the model proposed in Model Specification.
Table 1 The result of fitting Model Eq. (1) to the estimated smoking prevalence $\hat{x}$.

<table>
<thead>
<tr>
<th>Country</th>
<th>Universal parameters and Total Error ($E_2$)</th>
<th>Local parameters and local error ($E_{i,2}$ and $E_{i,1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b$</td>
<td>$\delta$  $E_2$</td>
</tr>
<tr>
<td></td>
<td>$a_1$ $x_1,0$ $u_{i,0}$ $u_{i,\infty}$ $E_{i,2}$ $E_{i,1}$</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1.049</td>
<td>0.9901 0.163</td>
</tr>
<tr>
<td>Canada</td>
<td>1.020</td>
<td>0.083 0.530 0.483 0.020 0.011</td>
</tr>
<tr>
<td>France</td>
<td>1.121</td>
<td>0.198 0.543 0.524 0.004 0.005</td>
</tr>
<tr>
<td>New Zealand</td>
<td>1.062</td>
<td>0.202 0.525 0.504 0.012 0.010</td>
</tr>
<tr>
<td>Sweden</td>
<td>1.076</td>
<td>0.077 0.555 0.503 0.015 0.009</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.976</td>
<td>0.079 0.513 0.478 0.060 0.018</td>
</tr>
<tr>
<td>United States</td>
<td>0.963</td>
<td>0.063 0.513 0.470 0.024 0.013</td>
</tr>
</tbody>
</table>

Table 2 Correlation between IDV, relative conformity $a$, average slope $s_x$, and peak year $t_{\text{max}}$.

<table>
<thead>
<tr>
<th>Country</th>
<th>7-country subset</th>
<th>25-country set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a$ $s_x$ $t_{\text{max}}$</td>
<td>$t_{\text{max}}$</td>
</tr>
<tr>
<td>IDV</td>
<td>-0.87 (0.011)</td>
<td>0.85 (0.015)</td>
</tr>
<tr>
<td></td>
<td>-0.76 (0.047)</td>
<td>-0.53 (0.006)</td>
</tr>
<tr>
<td>$a$</td>
<td>-0.92 (0.003)</td>
<td>0.88 (0.009)</td>
</tr>
<tr>
<td>$s_x$</td>
<td>-0.92 (0.003)</td>
<td>0.88 (0.009)</td>
</tr>
</tbody>
</table>

Correlations between IDV, $a$, $s_x$, and $t_{\text{max}}$ are recorded for the seven-country subset. Correlation between IDV and $t_{\text{max}}$ is recorded for the full set of 25 countries. $p$-values are in parentheses. All correlations are significant at the 95% confidence level.

Additional file 2 — Smoking prevalence and tobacco consumption data
CSV file containing four columns: country number as it appears in Table A.1 of the additional file Additional File 1.pdf, year ($t$), measurement ($x(t)$ or $c(t)$), and type of measurement (0 indicates a smoking prevalence measurement, while 1 indicates a cigarette consumption measurement).

Additional file 3 — Proxy data: articles published on the health effects of smoking
CSV which contains three columns: year ($t$), number of articles published in year $t$, and cumulative number of articles published up to and including year $t$ ($n(t)$).

Additional file 4 — Proxy data: US population
CSV containing two columns: year ($t$) and population $N_{\text{pop}}(t)$.

Additional file 5 — Matlab data files and simulation code
Matlab data files and simulation code used in preparation of this manuscript.
The influence of societal individualism on a century of tobacco use: modelling the prevalence of smoking

Appendices A and B

J.C. Lang, D.M. Abrams, and H. De Sterck

A Additional Tables and Figures

Table A.1: Summary of data on smoking prevalence $x$ and cigarette consumption $c$.

<table>
<thead>
<tr>
<th>No.</th>
<th>Country</th>
<th>Abbrev.</th>
<th>Smoking prevalence ($x$)</th>
<th>Cigarette consumption per person per day ($c$)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Obs. Period</td>
<td>No. of Obs.</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Australia</td>
<td>AUS</td>
<td>1964–2010</td>
<td>16</td>
<td>1920–2010</td>
</tr>
<tr>
<td>2</td>
<td>Austria</td>
<td>AUT</td>
<td>1972–2006</td>
<td>5</td>
<td>1923–2004</td>
</tr>
<tr>
<td>3</td>
<td>Belgium</td>
<td>BEL</td>
<td>1997–2008</td>
<td>4</td>
<td>1921–2011</td>
</tr>
<tr>
<td>4</td>
<td>Canada</td>
<td>CAN</td>
<td>1964–2011</td>
<td>29</td>
<td>1920–2010</td>
</tr>
<tr>
<td>5</td>
<td>Denmark</td>
<td>DNK</td>
<td>1970–2010</td>
<td>41</td>
<td>1920–2010</td>
</tr>
<tr>
<td>6</td>
<td>Finland</td>
<td>FIN</td>
<td>1978–2011</td>
<td>34</td>
<td>1920–2009</td>
</tr>
<tr>
<td>7</td>
<td>France</td>
<td>FRA</td>
<td>1960–2010</td>
<td>22</td>
<td>1900–2010</td>
</tr>
<tr>
<td>10</td>
<td>Iceland</td>
<td>ICE</td>
<td>1987–2012</td>
<td>26</td>
<td>1932–1995</td>
</tr>
<tr>
<td>13</td>
<td>Italy</td>
<td>ITA</td>
<td>1980–2012</td>
<td>23</td>
<td>1905–2010</td>
</tr>
<tr>
<td>14</td>
<td>Japan</td>
<td>JPN</td>
<td>1965–2011</td>
<td>47</td>
<td>1920–2007</td>
</tr>
<tr>
<td>17</td>
<td>Norway</td>
<td>NOR</td>
<td>1973–2012</td>
<td>40</td>
<td>1927–2011</td>
</tr>
<tr>
<td>20</td>
<td>Romania</td>
<td>ROM</td>
<td>–</td>
<td>0</td>
<td>1920–1995</td>
</tr>
<tr>
<td>21</td>
<td>Spain</td>
<td>SPA</td>
<td>1985–2011</td>
<td>11</td>
<td>1920–2010</td>
</tr>
<tr>
<td>23</td>
<td>Switzerland</td>
<td>CHE</td>
<td>1992–2007</td>
<td>4</td>
<td>1934–2009</td>
</tr>
<tr>
<td>24</td>
<td>United Kingdom</td>
<td>GBR</td>
<td>1960–2010</td>
<td>38</td>
<td>1905–2009</td>
</tr>
<tr>
<td>25</td>
<td>United States</td>
<td>USA</td>
<td>1965–2011</td>
<td>36</td>
<td>1920–2010</td>
</tr>
</tbody>
</table>
Table A.2: Result from Eq. (3) regression of smoking prevalence $x$ on cigarette consumption $c$.

<table>
<thead>
<tr>
<th>Country</th>
<th>$C \times 10^2$</th>
<th>$B \times 10^2$</th>
<th>$R^2$</th>
<th>$p$</th>
<th>$n_{obs}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>4.5 ± 1.3</td>
<td>−0.3 ± 8.8</td>
<td>0.80</td>
<td>$3.2 \times 10^{-6}$</td>
<td>16</td>
</tr>
<tr>
<td>Austria</td>
<td>0.0 ± 4.9</td>
<td>24.2 ± 32.4</td>
<td>0.00</td>
<td>0.99</td>
<td>4</td>
</tr>
<tr>
<td>Belgium</td>
<td>2.6 ± 20.3</td>
<td>13.0 ± 81.5</td>
<td>0.13</td>
<td>0.64</td>
<td>4</td>
</tr>
<tr>
<td>Canada</td>
<td>3.5 ± 0.5</td>
<td>6.3 ± 3.8</td>
<td>0.87</td>
<td>$3.0 \times 10^{-13}$</td>
<td>28</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.0 ± 9.2</td>
<td>40.5 ± 44.4</td>
<td>0.00</td>
<td>0.99</td>
<td>41</td>
</tr>
<tr>
<td>Finland</td>
<td>2.0 ± 0.7</td>
<td>15.8 ± 2.8</td>
<td>0.55</td>
<td>$1.0 \times 10^{-6}$</td>
<td>32</td>
</tr>
<tr>
<td>France</td>
<td>1.8 ± 0.5</td>
<td>19.1 ± 2.5</td>
<td>0.72</td>
<td>$6.3 \times 10^{-7}$</td>
<td>22</td>
</tr>
<tr>
<td>Greece</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Hungary</td>
<td>1.9 ± 1.6</td>
<td>17.4 ± 11.2</td>
<td>0.93</td>
<td>$3.5 \times 10^{-2}$</td>
<td>4</td>
</tr>
<tr>
<td>Iceland</td>
<td>4.9 ± 1.2</td>
<td>0.9 ± 7.0</td>
<td>0.93</td>
<td>$2.6 \times 10^{-5}$</td>
<td>9</td>
</tr>
<tr>
<td>Ireland</td>
<td>5.4 ± 1.1</td>
<td>−4.0 ± 7.4</td>
<td>0.93</td>
<td>$1.7 \times 10^{-6}$</td>
<td>11</td>
</tr>
<tr>
<td>Israel</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Italy</td>
<td>4.8 ± 2.5</td>
<td>−0.3 ± 13.2</td>
<td>0.47</td>
<td>$6.1 \times 10^{-4}$</td>
<td>21</td>
</tr>
<tr>
<td>Japan</td>
<td>1.3 ± 3.2</td>
<td>25.7 ± 27.2</td>
<td>0.02</td>
<td>0.43</td>
<td>43</td>
</tr>
<tr>
<td>Netherlands</td>
<td>4.8 ± 3.2</td>
<td>20.5 ± 15.0</td>
<td>0.32</td>
<td>$4.7 \times 10^{-3}$</td>
<td>23</td>
</tr>
<tr>
<td>New Zealand</td>
<td>2.0 ± 0.3</td>
<td>18.8 ± 1.4</td>
<td>0.86</td>
<td>$2.6 \times 10^{-12}$</td>
<td>27</td>
</tr>
<tr>
<td>Norway</td>
<td>−7.2 ± 4.3</td>
<td>50.1 ± 10.6</td>
<td>0.24</td>
<td>$1.6 \times 10^{-3}$</td>
<td>39</td>
</tr>
<tr>
<td>Poland</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Portugal</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Romania</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Spain</td>
<td>6.0 ± 6.2</td>
<td>−7.4 ± 41.7</td>
<td>0.38</td>
<td>$5.7 \times 10^{-2}$</td>
<td>10</td>
</tr>
<tr>
<td>Sweden</td>
<td>5.4 ± 0.6</td>
<td>4.3 ± 2.3</td>
<td>0.92</td>
<td>$1.7 \times 10^{-15}$</td>
<td>27</td>
</tr>
<tr>
<td>Switzerland</td>
<td>2.8 ± 5.6</td>
<td>7.2 ± 38.6</td>
<td>0.69</td>
<td>0.17</td>
<td>4</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>5.6 ± 0.7</td>
<td>1.6 ± 4.5</td>
<td>0.88</td>
<td>$5.3 \times 10^{-18}$</td>
<td>37</td>
</tr>
<tr>
<td>United States</td>
<td>3.6 ± 0.3</td>
<td>−0.1 ± 2.3</td>
<td>0.95</td>
<td>$1.1 \times 10^{-22}$</td>
<td>35</td>
</tr>
</tbody>
</table>

± indicates 95% confidence intervals. We report $R^2$ values for the linear regression of $x$ on $c$, the $p$-value of the correlation between $x$ and $c$, and the number of years for which both $x$ and $c$ measurements are available, $n_{obs}$.
Table A.3: Hofstede’s Individualism Index IDV and peak year in cigarette consumption ($t_{max}$)

<table>
<thead>
<tr>
<th>Country</th>
<th>IDV</th>
<th>Peak year ($t_{max}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>90</td>
<td>1974</td>
</tr>
<tr>
<td>Austria</td>
<td>55</td>
<td>1979</td>
</tr>
<tr>
<td>Belgium</td>
<td>75</td>
<td>1973</td>
</tr>
<tr>
<td>Canada</td>
<td>80</td>
<td>1976</td>
</tr>
<tr>
<td>Denmark</td>
<td>74</td>
<td>1976</td>
</tr>
<tr>
<td>Finland</td>
<td>63</td>
<td>1963</td>
</tr>
<tr>
<td>France</td>
<td>71</td>
<td>1985</td>
</tr>
<tr>
<td>Greece</td>
<td>35</td>
<td>1986</td>
</tr>
<tr>
<td>Hungary</td>
<td>80</td>
<td>1980</td>
</tr>
<tr>
<td>Iceland</td>
<td>60</td>
<td>1984</td>
</tr>
<tr>
<td>Ireland</td>
<td>70</td>
<td>1974</td>
</tr>
<tr>
<td>Israel</td>
<td>54</td>
<td>1974</td>
</tr>
<tr>
<td>Italy</td>
<td>76</td>
<td>1984</td>
</tr>
<tr>
<td>Japan</td>
<td>46</td>
<td>1977</td>
</tr>
<tr>
<td>Netherlands</td>
<td>80</td>
<td>1977</td>
</tr>
<tr>
<td>New Zealand</td>
<td>79</td>
<td>1975</td>
</tr>
<tr>
<td>Norway</td>
<td>69</td>
<td>2004</td>
</tr>
<tr>
<td>Poland</td>
<td>60</td>
<td>1991</td>
</tr>
<tr>
<td>Portugal</td>
<td>27</td>
<td>1994</td>
</tr>
<tr>
<td>Romania</td>
<td>30</td>
<td>1995</td>
</tr>
<tr>
<td>Spain</td>
<td>51</td>
<td>1985</td>
</tr>
<tr>
<td>Sweden</td>
<td>71</td>
<td>1976</td>
</tr>
<tr>
<td>Switzerland</td>
<td>68</td>
<td>1972</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>89</td>
<td>1973</td>
</tr>
<tr>
<td>United States</td>
<td>91</td>
<td>1963</td>
</tr>
</tbody>
</table>
Figure A.1: **Raw smoking prevalence and cigarette consumption data.** Raw smoking prevalence $x$ (blue asterisks, left axis) and raw cigarette consumption $c$ (black dots, right axis) versus time. Raw cigarette consumption data is given in grams per person per day. A single outlier for smoking prevalence ($x$) for the country of France (panel c) is denoted with a red asterisk.
Figure A.2: The result of fitting Eq. (1) to the estimated smoking prevalence \( \hat{x} \). Estimated smoking prevalence values \( \hat{x} \) are given by blue dots.
B Additional Remarks on Model Implications and Study Design

B.1 A Counterfactual Scenario

In this section we illustrate the effect size of individualism/collectivism on the dynamics of the smoking epidemic by considering a simple counterfactual scenario. Specifically, holding all other fitted parameters constant, we consider how the smoking epidemic in the United States might have evolved if the United States (IDV=91 and $a = 0.963$) were about 2\% less individualistic (IDV=89 and, using the slope from Fig. 4(a), $a = 0.974$). Fig. A.3 plots an estimate for the number of cigarettes smoked per year (in trillions) versus time. Integrating the difference between the number of cigarettes smoked per year versus time for the United States with fitted ($a = 0.963$, solid line) and counterfactual ($a = 0.974$, dashed line) relative conformity implies that, according to our model, if the United States had 2\% lower individualism during the 90 year period from 1920–2010 then there would have been approximately $7 \times 10^{12}$ fewer cigarettes smoked. This is equivalent to a 16\% decrease in the number of cigarettes smoked.

![Figure A.3: Solution to Eq. (1) for the United States with $a = 0.963$ (solid) and $a = 0.974$ (dashed). Parameters $x_0$, $b$, $u_0$, $u_\infty$, and $\delta$ are as reported for the United States in Table 1.](image)

The number of cigarettes smoked per year is estimated as follows. First, we observe that for year $t$ the number of cigarettes smoked per smoker per day is $C_a(t) = c(t)/x(t)$. Therefore, using Eq. (4) we find that the number of cigarettes smoked per smoker per year $C_a(t) = 365 \times C_a(t)$ can be bounded. For example, in the case of the United States, where $\hat{B} < 0$, we find that $C_a(t)$ is bounded by

$$1.02 \times 10^4 = 365 \times \hat{C}_{-1} \times \frac{365 \text{ days}}{\text{year}} \leq C_a(t) = \frac{1 - \hat{B}/\hat{x}(t)}{C} \times \frac{365 \text{ days}}{\text{year}} \leq \frac{1 - \hat{B}/\min \hat{x}(t)}{C} \times \frac{365 \text{ days}}{\text{year}} = 1.04 \times 10^4.$$

Since the lower and upper bounds are relatively tight, we estimate the number of cigarettes smoked per smoker per year to be the average of the lower and upper bounds

$$\bar{C}_a \approx \frac{2 - \hat{B}/\min \hat{x}(t)}{2C} \times \frac{365 \text{ days}}{\text{year}} \approx 1.0 \times 10^4.$$

\[1\] Assuming 1.002 cigarettes per gram, as in [39,40].
We cross-check this estimate with the direct estimate of $\bar{C}_a$ taken by averaging $c(t)/x(t)$ for all times where both measurements are available in the raw data (data shown in Fig. A.1). These two estimates agree to two significant figures. Finally, we estimate the number of cigarettes smoked per year to be

$$\dot{x}(t) \times N_{\text{pop}}(t) \times \bar{C}_a,$$

where $N_{\text{pop}}(t)$ is the total population at time $t$. The total population for the United States is taken from US census estimates [47,48] and is given CSV format in the additional file, Additional File 4.csv, which contains two columns: year ($t$) and population $N_{\text{pop}}(t)$.

We emphasize that in the counterfactual scenario described above we have only changed $a$ for the United States while keeping all other fitted parameters constant, merely to illustrate that the effect of small changes in $a$ in the model can be large. Therefore, the broad variation in the fitted $a$ across countries, as illustrated in Fig. 4(a), can indeed be expected to lead to a large effect size on cigarette consumption. Note, however, that the results from this counterfactual scenario do not imply that less individualism automatically means lower cigarette consumption, since countries with lower IDV (higher $a$) than the United States also tend to differ for other fitted parameters and quantities in the model, resulting in substantially different solutions to Eq. (1).

### B.2 Order of Model Development and Additional Analyses

The mathematical model was proposed and developed before the data sets were compiled. Following the specification of the model no modifications were made or required to produce the reported results. The correlation between $a$ and IDV was investigated after fitting the model to the data, and strong negative correlation was obtained as a confirmation of the mechanism proposed in the model. In a subsequent step, to further corroborate the hypothesis that societal individualism influences the temporal dynamics of smoking prevalence at the population level, the correlation between IDV and $t_{\text{max}}$ was also confirmed for the raw smoking data, independent of the mathematical model. No analysis was performed with additional variables. However, the sensitivity of the model to several assumptions was tested. For example, and as already mentioned, one alternative to the discounting function presented in Eq. (2) was tested: we assumed a step-function individual utility function that took value $u_0$ for $t_0 \leq t < t^*$ and $u_\infty$ for $t > t^*$. We also tested the model for various combinations of local and global parameters with both utility functions. For example, whereas in our model $\delta$ and $b$ were taken to be global parameters and $x_{i,0}$, $u_{i,0}$, $u_{i,\infty}$, and $a_i$ were taken to be local parameters, we also tested the cases where (a) $b$ was the only global parameter and $x_{i,0}$, $u_{i,0}$, $u_{i,\infty}$, $a_i$, and $\delta_i$ were taken to be local parameters, (b) $a$ and $b$ were taken to be global parameters and $x_{i,0}$, $u_{i,0}$, $u_{i,\infty}$, and $\delta_i$ were taken to be local parameters, and (c) $a$, $b$, and $\delta_i$ were taken to be global parameters and $x_{i,0}$, $u_{i,0}$, and $u_{i,\infty}$ were taken to be local parameters. These variations confirmed that our modelling procedure was robust, i.e. these variations all produced qualitatively similar results.