The invisible tax of free knowledge

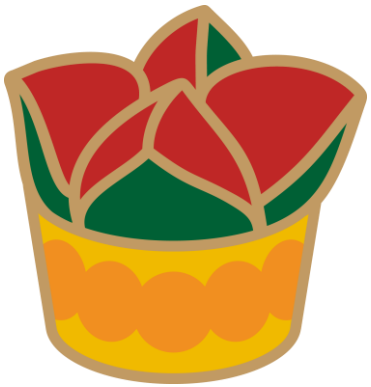
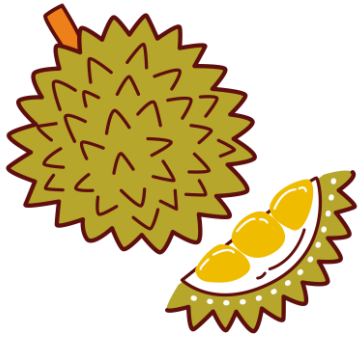
Evidence from the Wikimedia projects

Wikimania 2023

Singapore

Kiril Simeonovski



Introduction

- Free knowledge is commonly thought as, but in practice is not a pure public good.
  - while it is perfectly non-rivalrous, its non-excludability cannot be always achieved
  - in reality, it is an example of an impure public good.

Introduction

- Free knowledge is commonly thought as, but in practice is not a pure public good.
  - while it is perfectly non-rivalrous, its non-excludability cannot be always achieved
  - in reality, it is an example of an impure public good

- Excludability results from limitations of its access, and it is driven by economic, institutional and social factors.
  - less people can consume it
  - less people can produce it

} economic inefficiency

Research questions

- Why is free knowledge not a pure public good?
- What are the implications of the impurity?
- How to measure the implications of the impurity?
- How big are these implications across countries?
- What are the factors that contribute to the impurity?

Research outline

- Definition of pure public good vs impure public good.
- Model of free knowledge as a public good.
  - … relies on peer production in the Wikimedia movement
  - … uses Wikimedia content as a proxy of free knowledge
- Introduction of the concept of “invisible tax”.
- Calibration using country data from the Wikimedia projects.
- Identification of factors affecting the invisible tax rate.

Wikimedia vision

Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. That's what we're doing.

— Wales (2004)

Wikimedia vision

Imagine a world in which every single person on the planet is given free access to the sum of all human knowledge. That's what we're doing.

— Wales (2004)

The ultimate goal is to make the content on the Wikimedia projects a pure public good.

Related literature

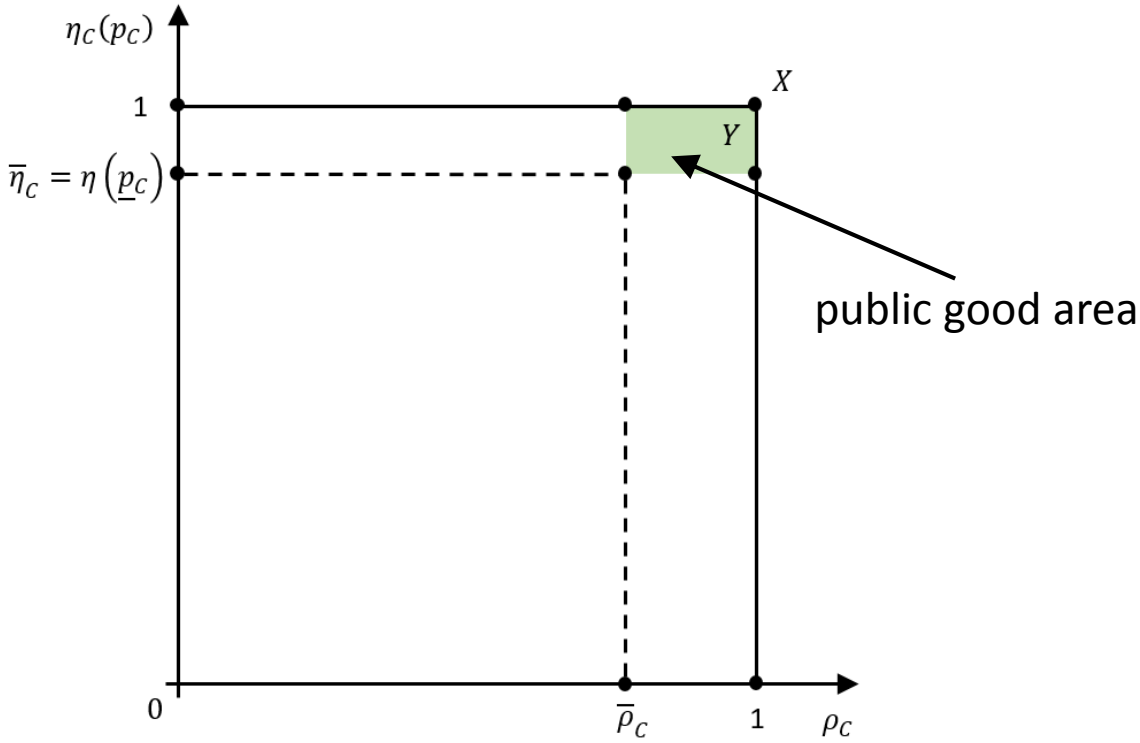
- Economic literature on modelling free knowledge is scarce, but there are papers on contribution to public goods.
  - Voluntary provision models (Cornes & Sandler 1985, 1994; Bergstrom et al. 1986; Epple & Romano 2003; Cornes & Hartley 2007, 2012; Burger & Kolstad, 2009; Kotchen 2009; Freundt & Lange 2021)
  - Models with pure altruism (Palfrey & Rosenthal 1984; Andreoni 1988; Fries et al. 1991)
  - Models with impure altruism (Cornes & Sandler 1984; Steinberg, 1987; Andreoni 1989, 1990)
  - Fairness and reciprocity (Rabin 1993; Fehr 2000, Fehr & Schmidt, 2006; Dufwenberg & Kirchsteiger 2004; Falk & Fischbacher 2006)
  - Social image and pro-social behaviour (Holländer 1990; Bénabou & Tirole 2006; Andreoni & Bernheim 2009; Ellingsen & Johannesson 2008, 2011; Bursztyn & Jensen 2017)
- Literature with insights from the Wikimedia projects (Zhang & Zhu 2011; Algan et al. 2013; Hergueux & Jacquemet 2015)

Pure vs impure good

- Every good in the economy G has two properties:
  - excludability $\eta \in [0; 1]$ ($\eta = 1$ denotes perfect non-excludability)
  - rivalry $\rho \in [0; 1]$ ($\rho = 1$ denotes perfect non-rivalry)
- A good G usually has a complementary good C so that its excludability $\eta = \eta(C, p)$ is an increasing function of the state of their complementary goods C and their price p.
- If C is a complementary good with no complementary goods and $\underline{p}_C$ is the lowest price that some individuals cannot afford to pay, then
  - $\overline{\eta}_C = \eta\left(\underline{p}_C\right)$ is the highest level of excludability at which there are individuals who cannot access the good
  - $\overline{\rho}_C$ is the highest level of rivalry at which there are individuals who cannot consume the good

Pure vs impure good

- Definition: A complementary good C is
  - a pure public good if $\eta_C = 1$ and $\rho_C = 1$
  - an impure public good if $\eta_C > \overline{\eta}_C$ and $\rho_C > \overline{\rho}_C$
  - a private good if $\eta_C \leq \overline{\eta}_C$ or $\rho_C \leq \overline{\rho}_C$



Pure vs impure good

- Definition: A public good G is
  - pure if $\rho = 1, p = 0$ and C is a **public** good
  - impure if $\rho = 1, p = 0$ and C is a **private** good

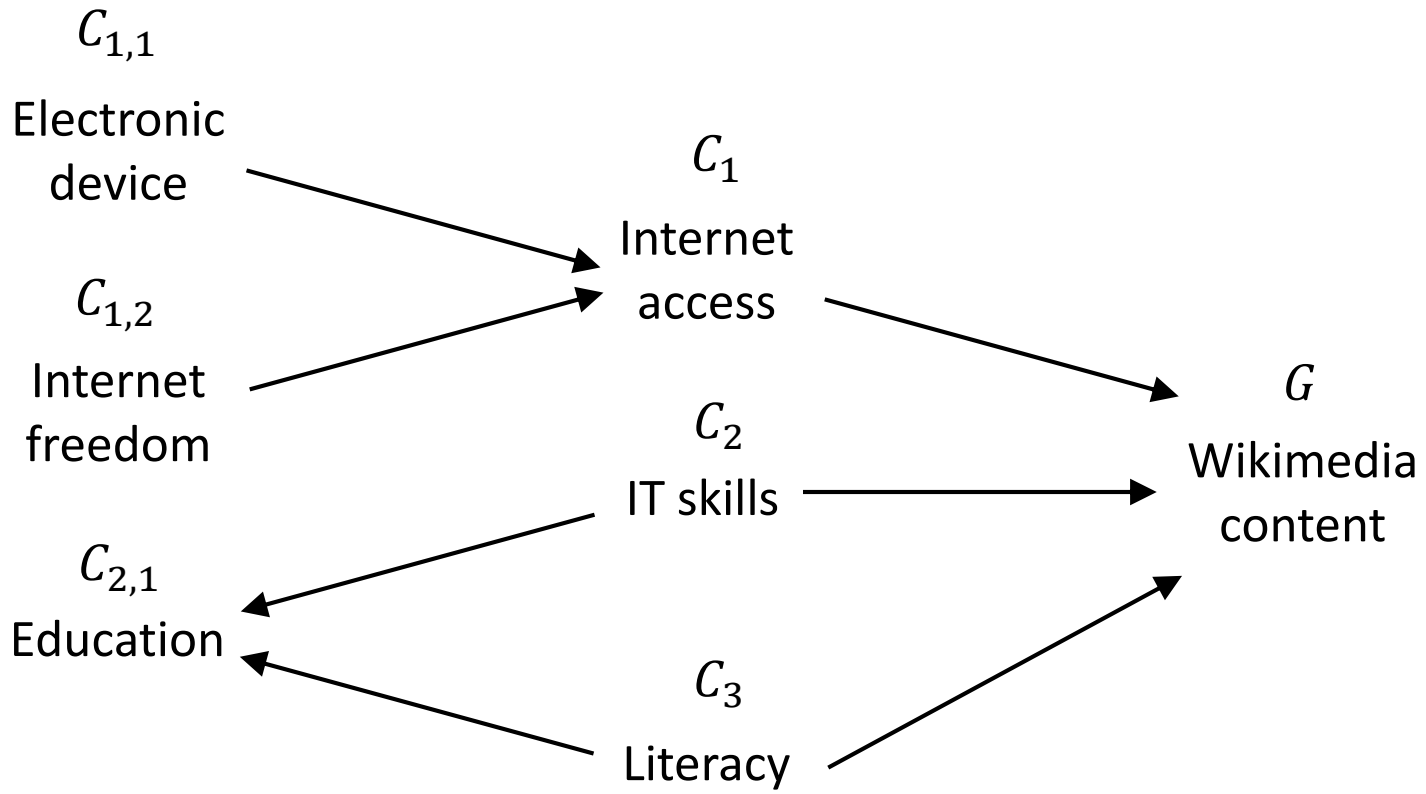
Proposition

A public good G is pure if and only if for each sequence of complementary goods $C_{j,k} = \{C_{j,k}\}_{k=1}^{n_1}$ from the vector $\mathbf{C} = \left(\{C_{1,k}\}_{k=1}^{n_1}, \dots, \{C_{m,k}\}_{k=1}^{n_m}\right)$ it holds that $C_k : Y \to \mathbb{R}$.

Implication: Individuals who cannot afford to pay for one in the network of complementary goods are unable to access the final good.

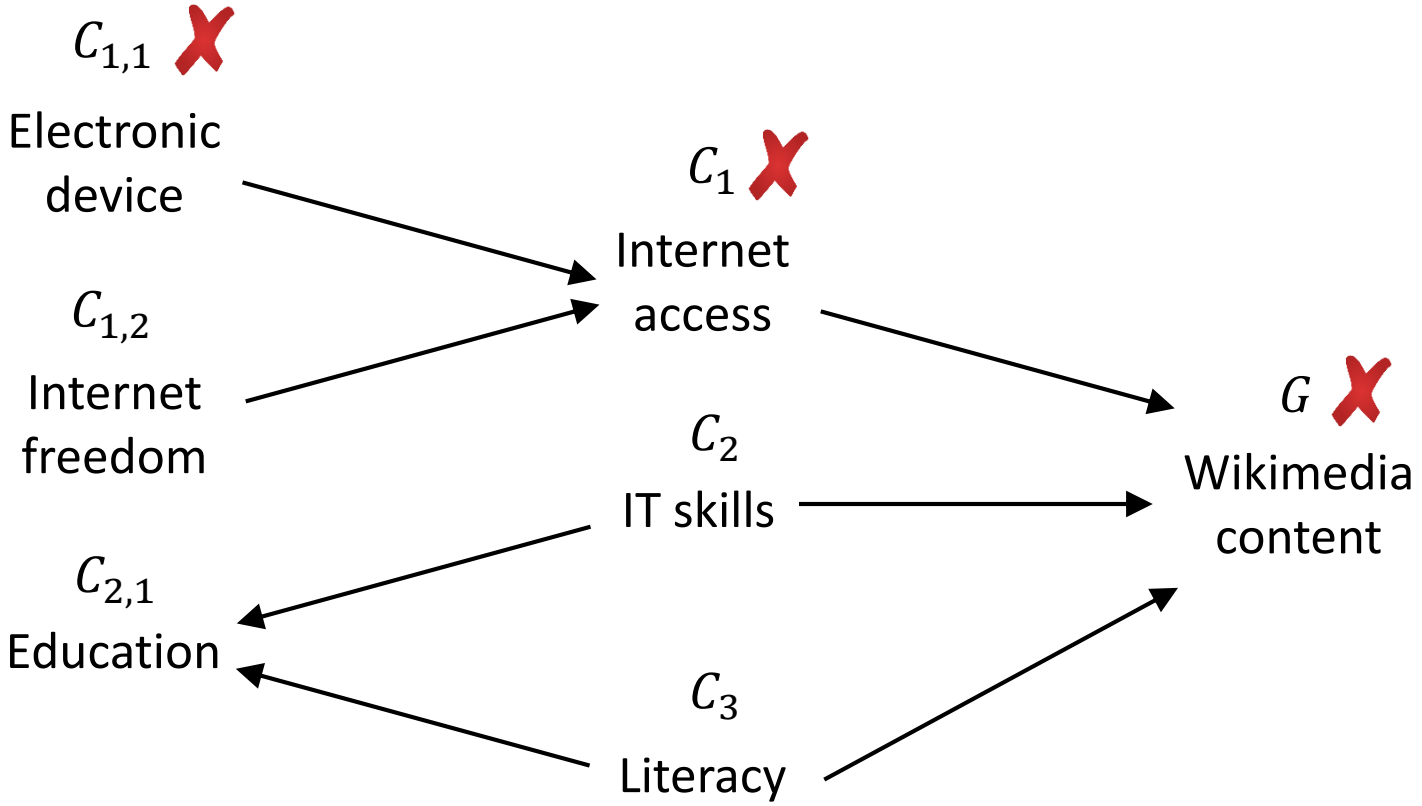
Pure vs impure good

Example



Pure vs impure good

Example



Economic environment

- There is a finite number of individuals and time is discrete and infinite, yet free knowledge is produced in a continuous time setting.
- Individuals spend their leisure time h_i on producing (w_i) and consuming free knowledge (v_i) where contribution time $w_i = w(\zeta_i, N, D)$ is an increasing and concave function of the altruism degree ζ_i, population size N and development level D.
- Definition: An individual can be:
  - … an altruist ($\zeta_i > 0$ and $w_i > 0$)
  - … an egoist ($\zeta_i = 0$ and $w_i = 0$)
- The population based on the altruism level can be decomposed as

$$N = M + R$$

where M is the share of individuals contributing to the free knowledge and R is the share of free riders.

Production of free knowledge

- Individual's contribution to the production of free knowledge is

$$g_i = \mu_i w_i$$

where μ_i is a non-negative productivity rate.

- Total amount of free knowledge produced in the economy is

$$G = G(T) = \sum_{i=1}^{N} \left(\int_0^T \mu_i(t) w_i(t)\, dt \right)$$

- Total free knowledge can be decomposed as

$$G = g_{i,t} + G_{-i,t} + G_{T-1}$$

where $G_{-i,t}$ is the free knowledge produced by all other individuals in the economy and G_{T-1} is the total amount of free knowledge produced in the previous periods.

Social interactions

- Free knowledge is produced in an environment with social interactions.
- Utility of social interactions can be expressed as

$$F_i = f\left(\sum_{j>i}^{N} \delta_{ij} F_j \underbrace{\Delta\mu_{ij} a_{ij}\left(w_i + w_j\right)}_{\Delta g_{ij}}\right)$$

where

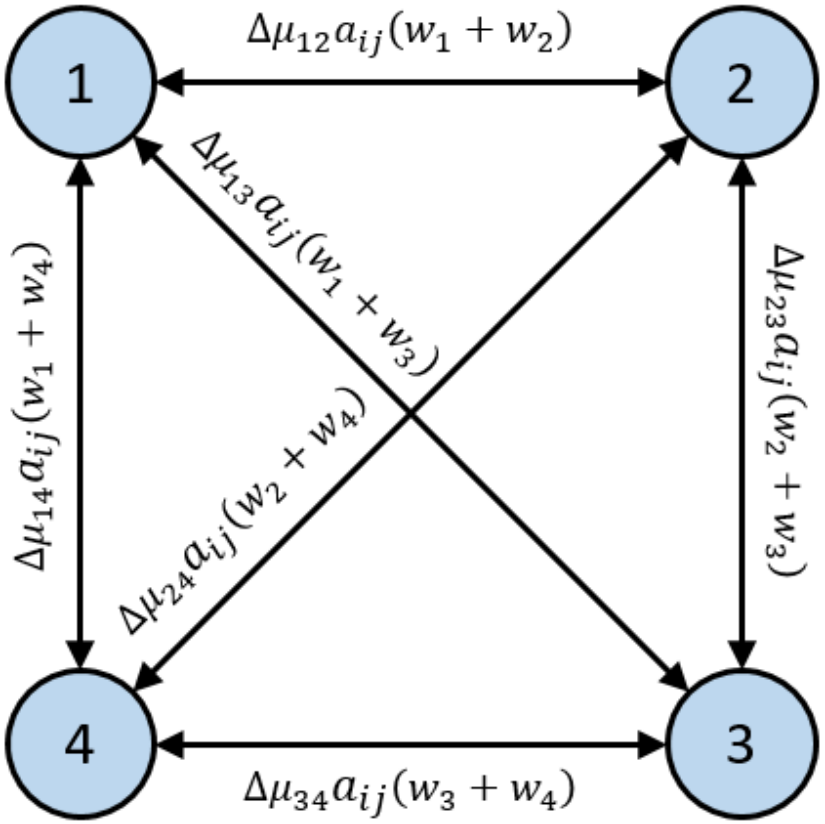
- $\delta_{ij} \in [0; 1]$ is the degree of interaction between individuals
- $\Delta\mu_{ij} = \mu_{ij}/\mu_i \in \mathbb{R}$ is the excess productivity from interaction
- $a_i = \sum_{j>i}^{N} a_{ij} \leq 1/2$ is the share of time devoted to interactions

- Total amount of free knowledge produced in the economy is

$$G = \sum_{i=1}^{N}\left(\int_0^T \mu_i(t) w_i(t) dt\right) + \int_0^T W(t)\left[\Delta\mu_i(t) a_i(t) + \Delta\mu_j(t) a_j(t)\right] dt$$

Social interactions

- A graph with four vertices and six directed edges.



Individual's utility problem

- Individual's problem is

$$\max_{w_{i,t}}\{u(v_{i,t}, G) + S(w_{i,t}, G, F_{i,t})\}$$

where $u(v_{i,t}, G)$ is the utility of consumption and $S(w_{i,t}, G, F_{i,t})$ is the social benefit of production.

- Marginal utilities:

  $u'_v > 0, S'_w > 0, u'_G > 0, S'_G < 0, u'_v > S'_w$ and $u'_G + S'_G > 0$

Proposition

The Nash equilibrium in the production of free knowledge is achieved when $w_{i,t} = 0$ and $G_t = 0$ if and only if $S(w_{i,t}, G_t) = 0$, as well as when $w_{i,t} > 0$ and $G_t > 0$ if and only if $S(w_{i,t}, G_t) > 0$.

Intuition: If a rational individual derives no benefit from doing something, she will opt to not do it.

Equilibrium characterisation

- Aggregate demand represents the aggregate marginal utilities

$$AD_t = U_t = \sum_{i=1}^{N} \left(\frac{du_{i,t}}{dG} + \frac{dS_{i,t}}{dG} \right)$$

- Aggregate supply is the total amount of free knowledge

$$AS_t = G = \left[\mu + a\left(\Delta\mu_i + \Delta\mu_j \right) \right] \int_0^T W(t)\, dt$$

- Definition: State at which decisions made by individuals $\mathbf{\Phi} = (\Phi_1, \dots, \Phi_n)$ on contribution time $\mathbf{w} = (w_1, \dots, w_n)$ and $\mathbf{F} = (F_1, \dots, F_n)$ maximise utility $\mathbf{u} + \mathbf{s} = (u_1 + s_1, \dots, u_m + s_m, u_{m+1}, \dots, u_n)$.

Theorem

The equilibrium exists and is generically unique.

Effect of excludability and rivalry

- The share of population with access to free knowledge is

$$Z = N - Q = (1 - q)N$$

where Z is the share of individuals with access to free knowledge, whereas $Q = (1 - q)$ is the share with no access to free knowledge.

- The excludability rate $q = q(\mathbf{p}, \boldsymbol{\rho}, \mathcal{P})$ is a function of the vectors or excess prices $\mathbf{p} = (\Delta p_1, \dots, \Delta p_n)$ and excess rivalries $\boldsymbol{\rho} = (\Delta \rho_1, \dots, \Delta \rho_n)$, as well as the degree of artificial excludability imposed by state $\mathcal{P}$.

- Aggregate demand is

$$AD_t^Z = U_t^Z = \sum_{i=1}^{Z} \left(\frac{du_{i,t}}{dG} + \frac{dS_{i,t}}{dG} \right) < U_t = AD_t$$

- Aggregate supply is

$$AS_t^Z = G^Z = (1 - q)\left[\mu + a\left(\Delta \mu_i + \Delta \mu_j\right)\right] \int_0^T W(t)\, dt < G = AS_t$$

Invisible tax of free knowledge

- Definition: The invisible tax reflects the lower supply of free knowledge as a result of excludability and rivalry, and it can be calculated as

$$\tau_t = \frac{G_t^Q}{G_t}$$

where $G_t^Q = G_t - G_t^Z$ is the lower supply of free knowledge.

Invisible tax of free knowledge

- Definition: The invisible tax reflects the lower supply of free knowledge as a result of excludability and rivalry, and it can be calculated as

$$\tau_t = \frac{G_t^Q}{G_t}$$

where $G_t^Q = G_t - G_t^Z$ is the lower supply of free knowledge.

- Why to call it an “invisible tax”?

Invisible tax of free knowledge

- Definition: The invisible tax reflects the lower supply of free knowledge as a result of excludability and rivalry, and it can be calculated as

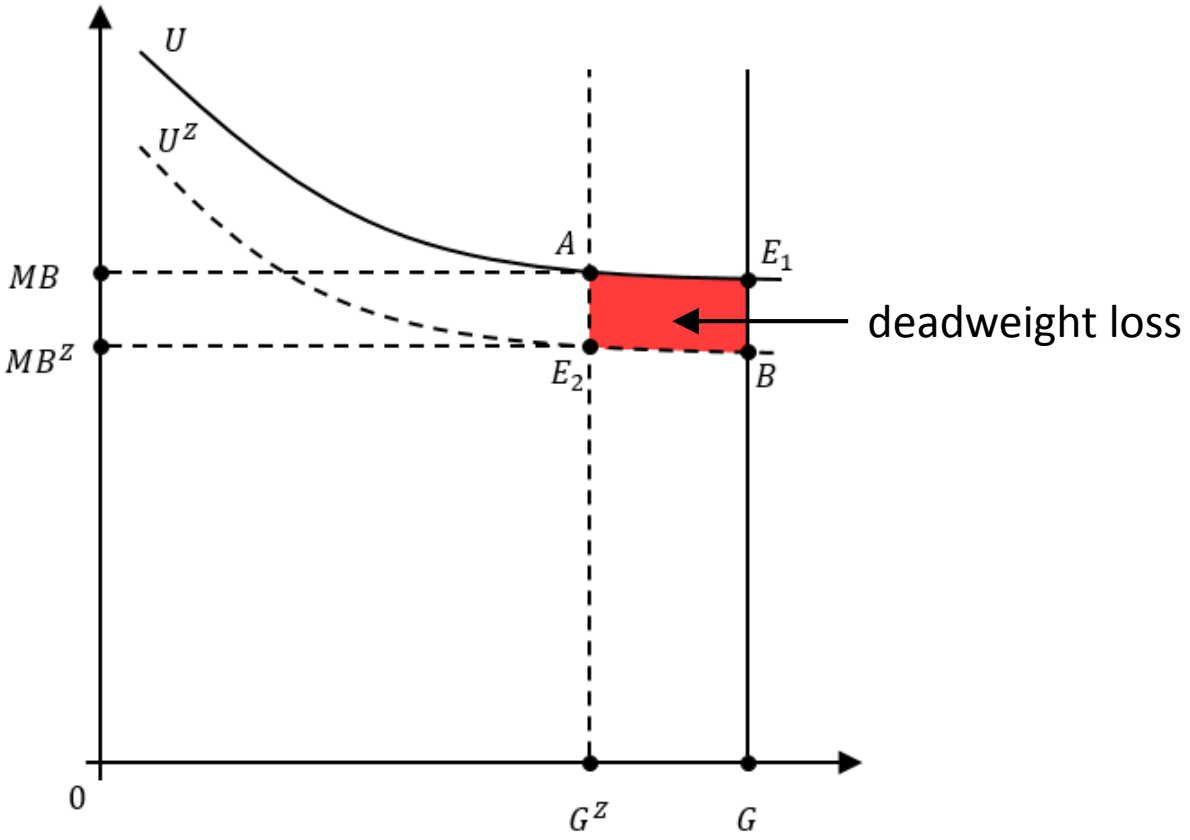
$$\tau_t = \frac{G_t^Q}{G_t}$$

where $G_t^Q = G_t - G_t^Z$ is the lower supply of free knowledge.

- Why to call it an “invisible tax”?
  - in public economics, a tax is an amount levied to support production and provision of public goods
  - in microeconomics, a tax is a source of economic inefficiency, which results in lower supply and demand (deadweight loss)
  - it is invisible because there is no monetary payment

Invisible tax of free knowledge

- Supply and demand shifts as a result of taxing free knowledge.



Invisible tax of free knowledge

Theorem

The deadweight loss of taxing free knowledge is the sum of utility functions of individuals with no access to free knowledge, that is

$$\mathcal{L}_t = \sum_{i=1}^{Q} \left[u\left(v_i, G_t^Q\right) + S\left(w_i, G_t^Q\right) \right]$$

Intuition: Those who do not have access to free knowledge cannot enjoy the benefits of consuming it and contributing to its production.

Calibration

- Data were obtained from the Wikimedia Foundation's databases.
  - Missing data on page edits for many countries, including Russia, China, Pakistan, Iran, Turkey and Thailand among others.
- Annual elasticities of page edits estimated with the quadratic regression

$$Page\ edits\ per\ capita_i = \alpha + \beta_1 Share\ of\ Internet\ users_i + \beta_2 Literacy\ rate_i + \beta_3 Literacy\ rate_i^2 + \varepsilon$$

- Aggregating page edits per country using the formula

$$\sum_{b=1}^{2} \sum_{l=1}^{L} \overline{Page\ edits}_b \cdot \overline{Number\ of\ editors}_{b,l}$$

- Average page edits per buckets (5 to 99 edits and 100 or more edits) and average number of editors (intervals of ten) were calculated using simple interval means and normalisation to the aggregate number of page edits.

Calibration

- Annual elasticities of page edits per capita:

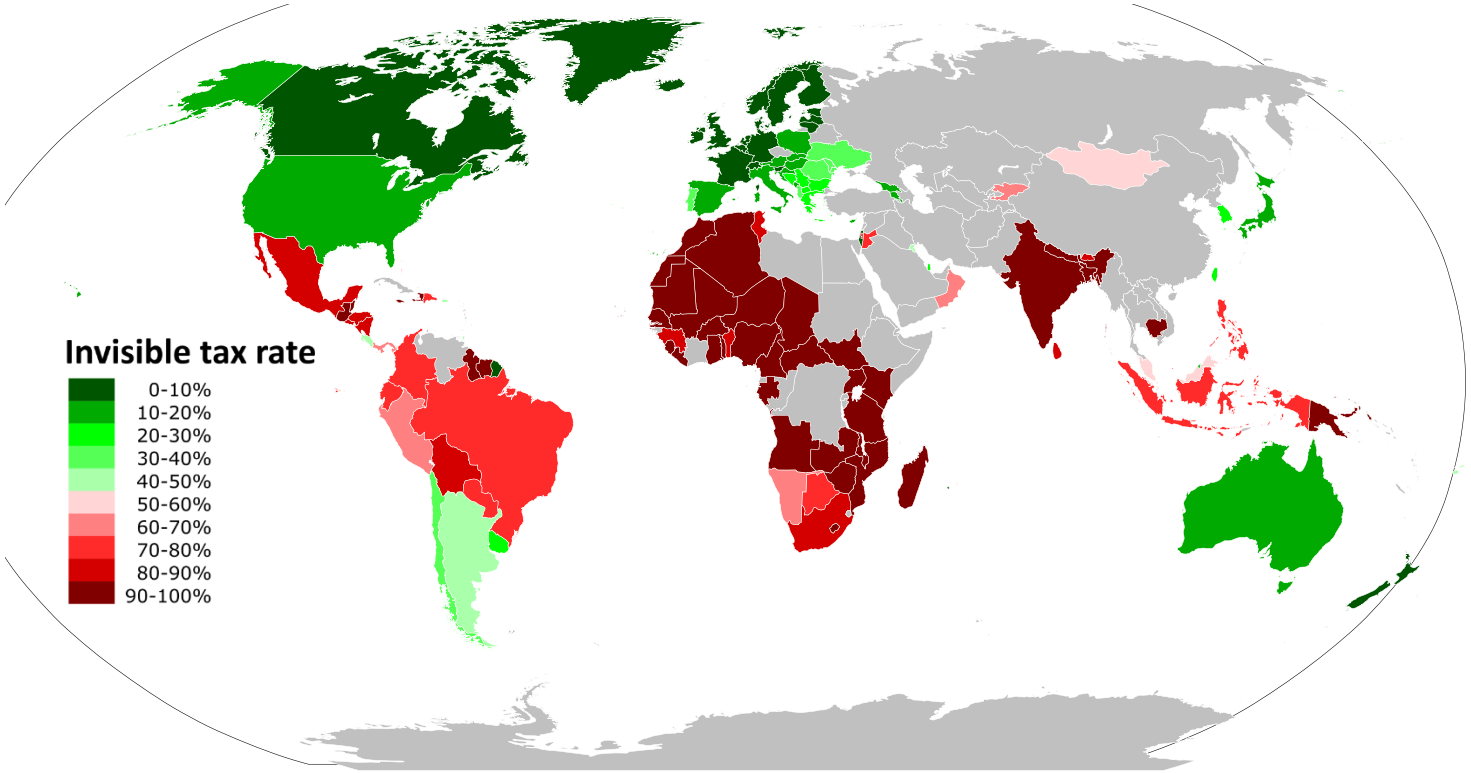
Variable	2021	2022
$Share\ of\ Internet\ users$	0.3282**	0.2960***
	(0.1438)	(0.1096)
$Literacy\ rate$	-1.5652***	-1.8668***
	(0.3959)	(0.4200)
$Literacy\ rate^{2}$	1.2425***	1.4713***
	(0.3699)	(0.3374)
Intercept	0.3507***	0.4394***
	(0.1256)	(0.1377)
Number of observations	144	144
Vertex	63.0%	63.4%

Notes: Robust standard errors are reported in parentheses. Symbols ***, ** and * denote statistical significance at the level of 1%, 5% and 10%, respectively.

- Using the elasticities to calculate the potential maximum of edits made and Wikipedia articles created.

Results

- Invisible tax of free knowledge in 2022 was 55.5% globally (56.9% in 2021).
  - Lowest rates: Luxembourg (0.3%), Norway (0.6%) and Finland (1.4%)
  - Highest rates: Malawi (99.8%), Chad (99.7%) and Lesotho (99.6%)
  - Global South (77.2%) vs Global North (14.6%)



Factors of excludability

- Digital divide.
  - in 2022, the average share of Internet users was 53.4% in Global South and 88.8% in Global North
- Net neutrality vs zero-rating.
  - zero-rating has positive economic effects for consumers (Rogerson 2016; Somogyi 2016; Krämer & Peitz 2018; Jeitschko et al. 2019)
  - Wikipedia Zero as an attempt to reduce excludability.
- Censorship.
  - blocking content and prosecuting editors (China, Russia, Turkey, Iran, Pakistan, Myanmar, Syria, Venezuela, Belarus and Saudi Arabia)
  - disputes related to single articles (United Kingdom, Australia, France and Germany)

Future research

- Re-calibration of the model with more precise and more granular data.
  - Number of edits vs bucketed number of editors.
  - Number of edits across languages and countries vs aggregated number of edits across countries.
- Further elaboration of the model's components and its extensions.
  - Modelling marginal utility functions across languages and countries.
  - Estimating and forecasting the contribution time w_i.
  - Studying social interactions and endogenising productivity rate μ_i.
- Conducting economic experiments.
  - Natural experiments to study the effect of reforms and censorship.
  - Online experiments to study behavior and preferences.

References

- Algan, Y., Benkler, J., Morell, M. F. & Hergueux, J. (2013). "Cooperation in Peer-Production Economy: Experimental Evidence from Wikipedia." *Sciences Po publications*.
- Andreoni, J. (1988). "Privately Provided Public Goods in a Large Economy: The Limits of Altruism." *Journal of Public Economics*, 35(1): 57–73.
- Andreoni, J. (1989). "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy*, 97(6): 1447–1458.
- Andreoni, J. (1990). "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *The Economic Journal*, 100(401): 464–477.
- Andreoni, J. & Bernheim, B. D. (2009). "Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica*, 77(5): 1607–1636.
- Bénabou, R. & Tirole, J. (2006). "Incentives and Prosocial Behavior." *American Economic Review*, 96(5): 1652–1678.
- Bergstrom, T. C., Blume, L. & Varian, H. (1986). "On the private provision of public goods." *Journal of Public Economics*, 29(1): 25–49.
- Burger, N. E. & Kolstad, C. D. (2009). "Voluntary Public Goods Provision, Coalition Formation, and Uncertainty." *NBER Working Papers*, 15543, National Bureau of Economic Research.
- Bursztyn, L. & Jensen, R. (2017). "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure". *Annual Review of Economics*, 9:131–153.
- Cornes, R. & Hartley, R. (2007). "Aggregative Public Good Games." *Journal of Public Economic Theory*, 9(2): 201–219.
- Cornes, R. & Hartley, R. (2012). "Fully aggregative games." *Economic Letters*, 116(3): 631–633.
- Cornes, R. & Sandler, T. (1984). "Easy Riders, Joint Production, and Public Goods." *The Economic Journal*, 94(375): 580–598.
- Cornes, R. & Sandler, T. (1985). "The simple analytics of pure public good provision." *Economica*, 52: 103–116.
- Cornes, R. & Sandler, T. (1994). "The Comparative Static Properties of the Impure Public Good Model." *Journal of Public Economics*, 54(3): 403–421.
- Dufwenberg, M. & Kirchsteiger, G. (2004). "A theory of sequential reciprocity. Games and Economic Behavior." 47(2): 268–298.
- Ellingsen, T. & Johannesson, M. (2008). "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review*, 98(3): 990–1008.
- Ellingsen, T. & Johannesson, M. (2011). "Conspicuous generosity." *Journal of Public Economics*, 95(9–10): 1131–1143.
- Epple, D. & Romano, R. (2003). "Collective Choice and Voluntary Provision of Public Goods." *International Economic Review*, 44(2): 545–572.
- Falk, A. & Fischbacher, U. (2006). "A theory of reciprocity. Games and Economic Behavior", 54(2): 293–315.
- Fehr, E. (2000). "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives*, 14(3): 159–181.
- Fehr, E. & Schmidt, K. M. (2006). "The Economics of Fairness, Reciprocity and Altruiusm – Experimental Evidence and New Theories." In Kolm, S. C. & Ythier, J. M. (eds.). *Handbook of the economics of giving, altruism and reciprocity*, 1: 615–691.
- Fries, T. L.., Golding, E. & Romano, R. (1991). "Private Provision of Public Goods and the Failure of the Neutrality Property in Large Finite Economies." *International Economic Review*, 32(1): 147–157.
- Freundt, J. & Lange, A. (2021). "On the voluntary provision of public goods under risk." *Journal of Behavioral and Experimental Economics,* 93, 101727.
- Hergueux, J. & Jacquemet, N. (2015). "Social preferences in the online laboratory: a randomized experiment." *Experimental Economics*, 18(2): 251–283.
- Holländer, H. (1990). "A Social Exchange Approach to Voluntary Cooperation." *American Economic Review*, 80(5): 1157–1167.
- Jeitschko, T. D.; Kim, S. J. & Yankelevich, A. (2019). "Zero-Rating and Vertical Content Foreclosure." *Information Economics and Policy*, 55, 100899.
- Kotchen, M. J. (2009). "Voluntary Provision of Public Goods for Bads: A Theory of Environmental Offsets." *The Economic Journal*, 119(537): 883–899.
- Krämer, J. & Peitz, M. (2018). "A fresh look at zero-rating." *Tellecommunication Policy*, 42(7): 501–513.
- Palfrey, T. R. & Rosenthal, H. (1984). "Participation and the Provision of Discrete Public Goods: A Strategic Analysis." *Journal of Public Economics*, 24(2): 171–193.
- Rogerson, W. P. (2016). "The Economics Of Data Caps And Free Data Services In Mobile Broadband." *CTIA report*.
- Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5): 1281–1302.
- Somogyi, R. (2016). "The Economics of Zero-Rating and Net Neutrality." *LIDAM Discussion Papers CORE 2016047*. Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).
- Steinberg, R. (1987). "Voluntary Donations and Public Expenditures in a Federalist System." *American Economic Review*, 77(1): 24–36.
- Zhang, M. X. & Zhu, F. (2011). "Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia." *American Economic Review*, 101(4): 1601–1615**.**