

Can we scale open-source development by removing the need for iconic and self-sacrificing leaders?

Simon Carbonnelle

ISP seminar

2nd of October 2019

Open-source

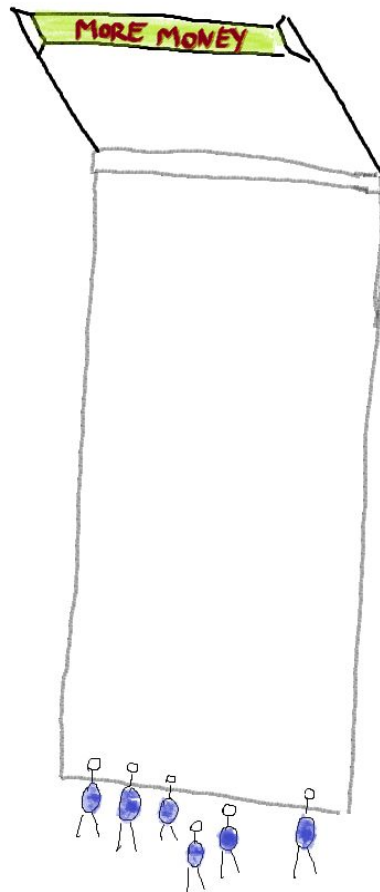
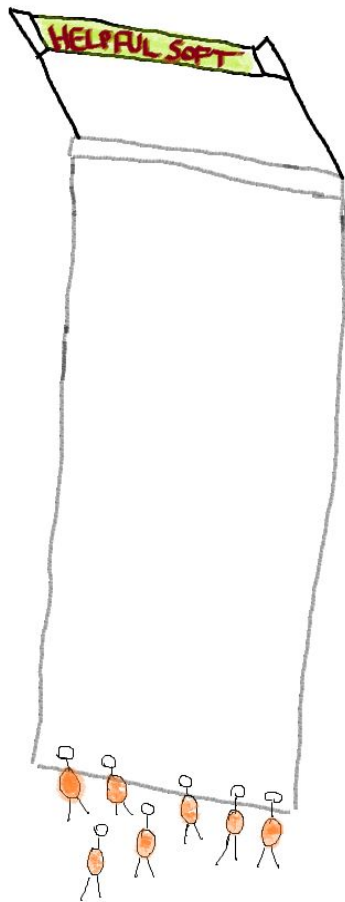
Users have the rights to study, change, and distribute the software to anyone and for any purpose.

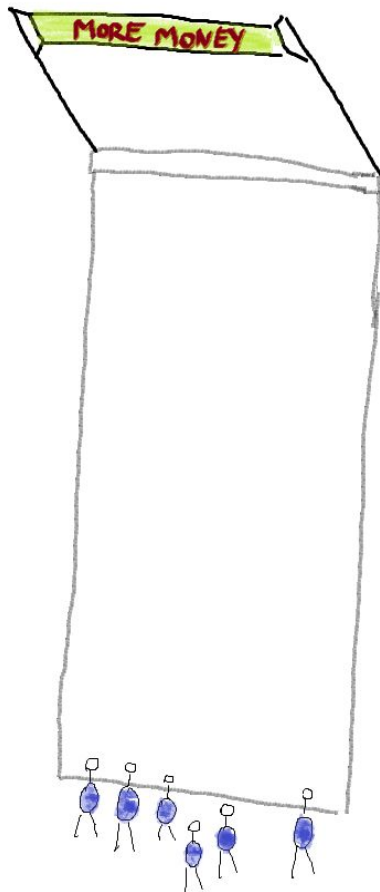
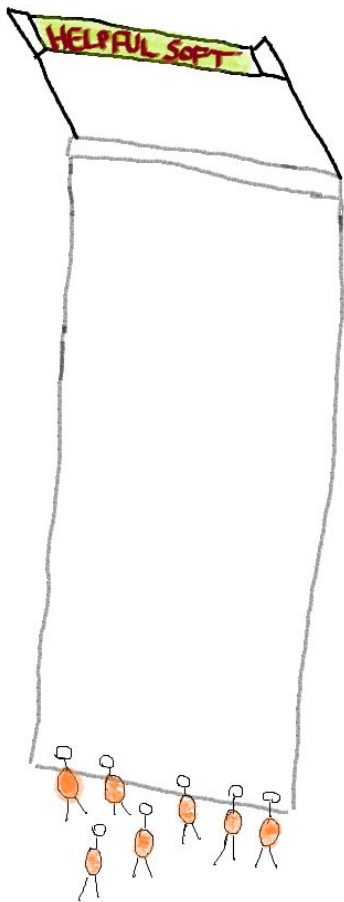
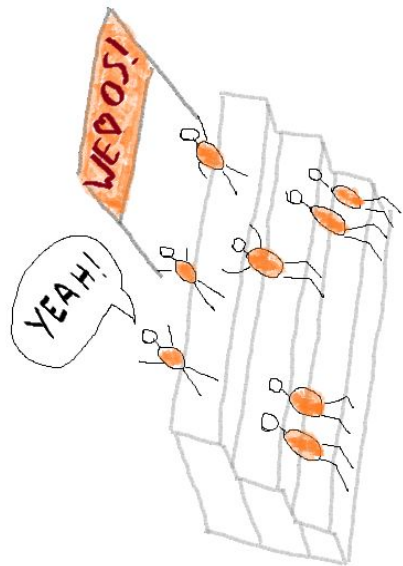


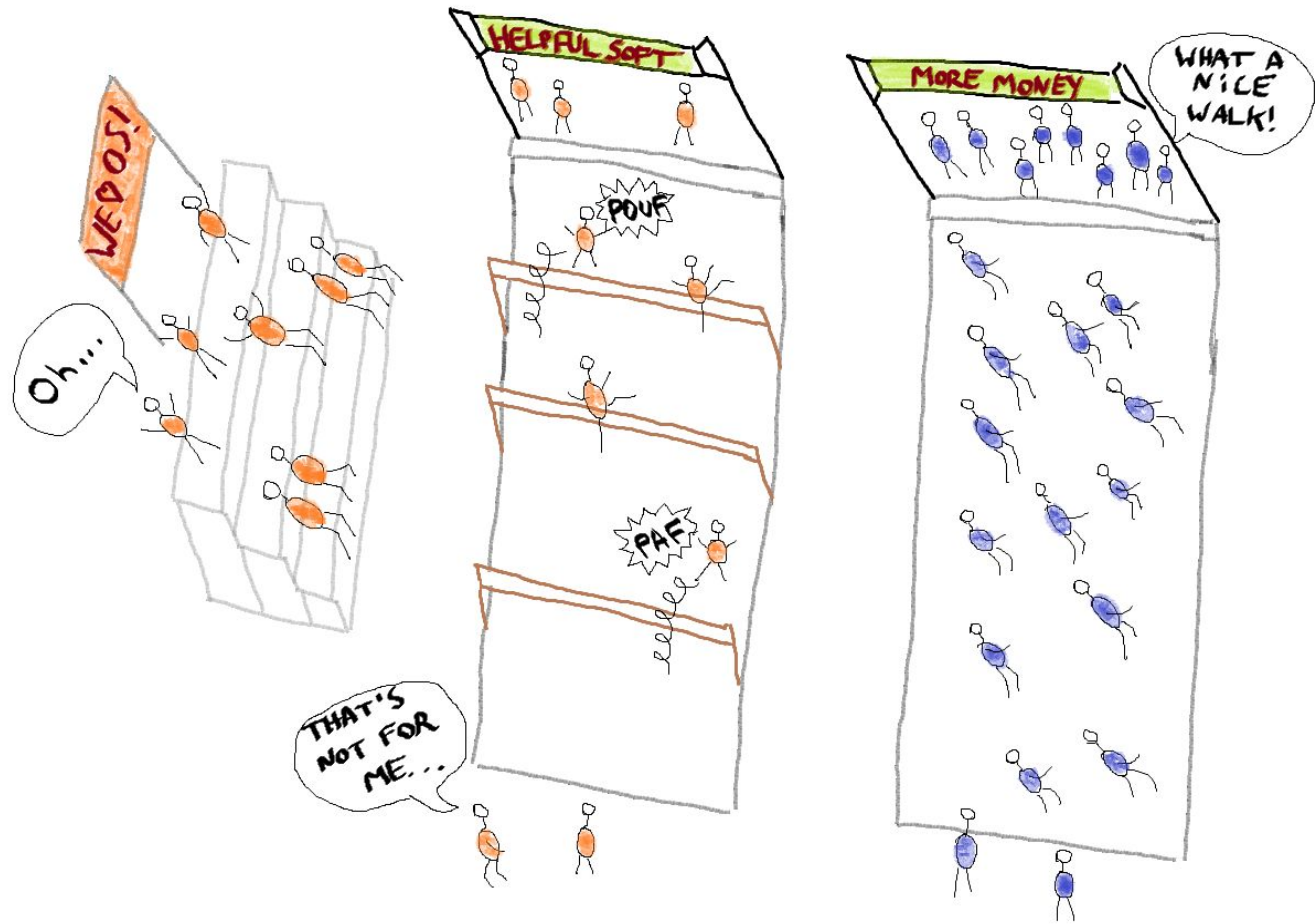
Proprietary

The software's publisher or another person retains intellectual property rights.









1. Understand the problems of open-source
2. Explore how we can rebalance forces

Why do individuals participate?

Organizational culture: the vision/values and social aspects of an organization

Open-source attracts members mainly through the cultural aspect of work.

Private sector attracts members mainly through salaries/compensations.

Which poses huge **constraints** for open-source:

- **Vision alignment or politics (Problem 1)**
The organization needs a transparent and shared vision
- **Quality of social interactions (Problem 2)**
Interactions between members need to be positive and stimulating

Vision alignment or politics

Examples from **open-source**

Fork off! FFmpeg project leader quits, says he's had enough with these forking AV libraries

Won't have Niedermayer to kick around anymore

By Neil McAllister in San Francisco 5 Aug 2015 at 07:01

47 

SHARE ▼

“The FFmpeg development model was fairly simple: the biggest FFmpeg contributor (Michael Niedermayer, see stats) was the **"self-proclaimed" maintainer.**”

“**The main complaints** I saw mentioned and remember from that time **were about the project leader.** It was all about "I do and commit whatever I want to", "I don't care about cosmetics", "I don't want to discuss with you in real life, mails are good enough". ”

Result: Community splitted into two forks... frustrating contributors and users

Vision alignment or politics

Examples from **open-source**

Why Neovim is Better than Vim

15 Jan 2015

“The codebase is atrocious. The plugin API is cumbersome and restrictive. The dev community is apathetic. The benevolent dictator is averse to change.”

“Speaking of Bram Moolenaar: **His merge criteria are inscrutable.** Some patches he ignores. Some, he attacks. Others, he merges.”

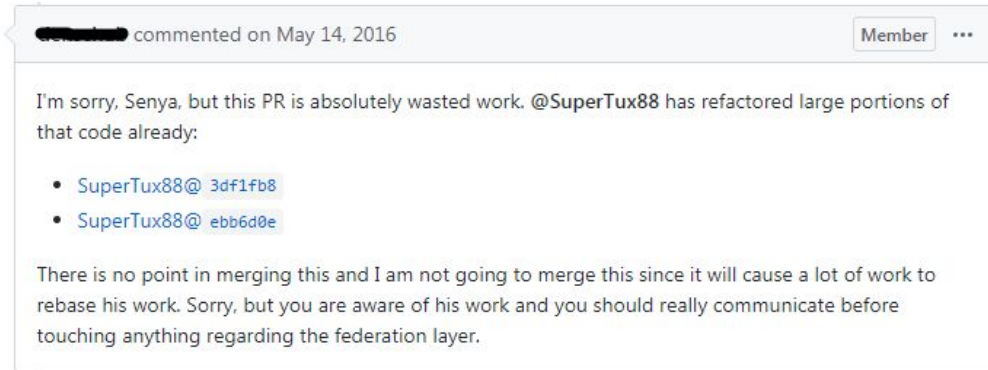
Result: Community splitted into two forks...

Quality of social interactions

Examples from **open-source**

Project management at diaspora* - or: what the heck am I doing here?

■ Governance



“The problem is that **the way you talk to people drive contributors away**. This is my case. And I assume that this was also the case of Jason.

You may be doing fantastic technical work, I won't judge that. But in the end, it remains that non-ore devs have to work with you to contribute. And this is my opinion on the matter : **working with you is not fun**. And this is a problem because **working for a free projet has to be fun to attract new contributors.**”

Vision alignment + social interactions

Examples from **open-source**

After governance breakdown, Node.js leaders fight for its survival

The tumultuous events this week in the Node.js community were a long time coming. Insiders tell ZDNet how the community is trying to repair — and prepare for what's to come.



By Zack Whittaker | August 25, 2017 -- 22:00 GMT (23:00 BST) | Topic: [Open Source](#)

"**Rod's behavior** was non-collaborative, consistent, and unapologetic," said Williams

The resignations followed a single event -- a **vote that failed to remove** a former director, a longstanding member of the community, from the leadership group.

Result: a **third of the technical steering committee quits**, a new fork is launched with core members

"I'm confident that **the individuals who voted on this didn't quite realize** how big of a thing this was going to be"

Vision alignment or politics

Examples from **private sector**

Inside Google's Civil War

Some employees say Google is losing touch with its "Don't be evil" motto. What happens when an empowered tech workforce rebels?

By [Beth Kowitt](#)

May 17, 2019

February 2018 PROJECT MAVEN LEAKS: Most employees **learn for the first time** that the Pentagon was using the company's A.I. to analyze drone footage.

August 2018 DRAGONFLY LEAKS: The Intercept reports that Google is working on a censored search engine in China; **it's the first time most employees are hearing of the project.**

Results: "Jack Poulson says he was the **sixth or seventh employee** to cite Dragonfly as a reason for quitting."

Out of 100.000 employees!!!

Vision alignment or politics

Examples from **private sector**

The New York Times

*Dozens at Facebook Unite
to Challenge Its
'Intolerant' Liberal Culture*

“The activity is a rare sign of organized dissent within Facebook”

1. Understand why open-source is losing



2. Explore how we can rebalance forces

How to make use of the available workforce?

Open-source contributors are mainly hobbyists, and demand flexible and lightweight workloads...

... but **full-timers** are needed in practice when an organization scales.

(**Problem 3**)

Flexible and lightweight workloads

Examples from **open-source**

Felix Krause

Scaling open source communities

Jan 31, 2017 |  [Edit on GitHub](#)

Not too long ago I started an open source project called fastlane. Just a month after publishing, it had 1000+ stars on GitHub and was beginning to get used by lots of serious tech companies around the world. **Soon I was the sole maintainer of a project getting 10+ PRs/day, and spending 8+hrs/day reviewing PRs and replying to questions on GitHub.**

Flexible and lightweight workloads

Examples from **open-source**

The core team – a bottleneck of the diaspora* development

■ Governance

“People who maintain the project – the coreteam – these are who are responsible for **planning, QA, code review, releases, they are the bottleneck.**”

“So **there are enough people who write code.** But it gets painfully long to get the code reviewed and merged.”

Flexible and lightweight workloads

Examples from **open-source**

[FFmpeg-devel] FFmpeg's future and resigning as leader

Michael Niedermayer [michael at niedermayer.cc](mailto:michael@niedermayer.cc)

Fri Jul 31 15:53:04 CEST 2015

“Indeed i fully admit **the work and pressure caused by the merges** is a main reason for my resignation.”

Flexible and lightweight workloads

Examples from **open-source**

Open source sustainability

Open source maintainers are exhausted and rarely paid. A new generation wants to change the economics.

“Working day jobs, maintainers today can struggle to find the time to fix critical bugs, all the while facing incessant demands from users requesting free support on GitHub. **Maintainer burnout is a monstrous challenge.**”

A complete report: *“Roads and Bridges: The Unseen Labor Behind Our Digital Infrastructure”*

1. Understand why open-source is losing



2. Explore how we can rebalance forces

Current practical solutions

Elaborated political systems

How Do Open Source Communities Govern Themselves?

Submitted by rfay on Mon, 2012-03-05 20:41

- Technical steering committee
- Community committee/ Community Council
- Elections (representative democracy)
- Constitution
- Code of Conduct/ Social Contracts

Current practical solutions

Funding



“open your finances to your community”



“The best way for artists and creators to get sustainable income and connect with their fans”

Popular “theoretical” solutions

Towards radically different forms of organization

- Decentralization, DAO’s (blockchain-based)

“An organization represented by rules encoded as a computer program that is transparent, controlled by shareholders and not influenced by a central government.”



- Horizontal organizations, holacracy, sociocracy, ...

“An organizational structure with few or no levels of middle management between staff and executives”



Not just about open-source!

Group decision making



From theory to practice?

My best guess

Tools that **increase our ability to cooperate on intellectual tasks at scale** are a crucial ingredient.

Getting better at thinking together

How? We can get there by augmenting our communication protocols with:

- 1) **Structure:** the relation between ideas/pieces of information is explicit
- 2) **Value-based feeds:** participants can share the value they confer to content

Structure

Examples

Doodle Pricing Help English ▾

Sign up

Log in

Create a Doodle

ICTEAM Spring Social Event

by Nathalie Ponet • 6 months ago • Print

	Je participe / I will be there	Je ne serai pas là / I won't be there	Je veux bien aider à la mise en place / I can help before the event	Je peux aider au rangement / I can help after the event
140 participants	✓ 117	✓ 22	✓ 18	✓ 18
<input type="text" value="Enter your name"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Itamar Levi	✓			
Massi	✓			
Arka Halder	✓			
Thomas Walewyns				
Sébastien Lugan	✓			
Danielle Vanhoenac...	✓			
François Aubry	✓			
Nikita Doikov	✓			
sophea chhun	✓		✓	✓

Arranging events through e-mail is a pain...

... Doodle solves the problem by offering a **structured** way to communicate around this task

(Framadate is a nice open-source alternative)

Structure Examples

The screenshot displays a Trello board for 'Chou Force Objectives'. The board is organized into several columns, each representing a different stage of task completion:

- Epics:** Contains high-level goals such as 'Launch V1 of Octalysis Prime Post-Island', 'Increase Q2 MRR by 60% to \$1,686.4', and 'Reduce Churn by 20% to 12%'.
- Inbox:** Lists tasks like 'Get a video testimonial from Willow about Octalysis Prime' and 'Answer questions on Quora to build a presence and link back to Ykc.com'.
- Prioritized:** Features tasks with priority indicators, such as 'Live AMA on Fb (with optional live video for the best questions)' and 'Create the 21-day Course that costs \$21'.
- This week:** Shows tasks scheduled for the current week, including 'Create the funnel for the email capture to Udemery course' and 'Set up the \$1 signup challenge after the Udemery course funnel'.
- Under discussion:** Contains tasks that are currently being discussed, such as 'Need to upgrade PHP on octalysisprime.com' and 'Integrate chouforce.com with our Square Space account'.
- Doing:** Lists tasks that are currently in progress, including 'Create all missing videos and get Post-Island (test version) launched in community' and 'Rewrite Onboarding Automation with more epic meaning, curiosity, and scarcity'.

Each card includes a title, a progress bar, a due date, and a count of comments or attachments. The board also shows a total count of cards for each column and a background image of a beach scene.

Value-based feeds

Examples

Reddit (Machine Learning community)

The screenshot displays the Reddit interface for the r/MachineLearning community. The main content area lists several posts, including a discussion about 'Machine Learning - WAYR (What Are You Reading) - Week 71' and a project about 'Using Machine Learning to detect handguns'. A green circle highlights the number '238' next to the second post. On the right, the 'Community Details' sidebar shows 782k members and 2.2k online users, with a red circle around these numbers. The 'CREATE POST' button is highlighted in blue.

Post Title	Engagement
[D] Machine Learning - WAYR (What Are You Reading) - Week 71	6
[D] AMA: I'm Dr. Genevieve Patterson - cofounder and Chief Scientist at TRASH, a new app that uses computer vision and computational photography to intelligently edit together and set to music any videos you upload. Ask me anything!	81
[N] UC Berkeley's CS 285: Deep Reinforcement Learning	0
Looking for tips on how to implement big data and data science programs in your organization? Check out this handy guide.	0
[R] Neural Ordinary Differential Equations: the current landscape	0
[P] Using Machine Learning to detect handguns being held by a person, not in holster - SmarterEveryDay [video 12:55]	20
[P] I scraped career statistics of Tour de France starters and used an LSTM network to predict their end of career	32
[R] On Model Stability as a Function of Random Seed	2
[R] Class Feature Pyramids for Video Explanation	0
Artificial Design: Modeling Artificial Super Intelligence with Extended General Relativity and Universal Darwinism via Geometrization for Universal Design Automation	51
[D] First time to NeurIPS	1
[D] What are your favorite research blogs that discusses advanced ML topics ?	39
[Project] Implementing Improvements to Hypergradient optimizers	2
TinyBERT: Distilling BERT for Natural Language Understanding	10

Value-based feeds

Examples

Reddit (Machine Learning community)

- ↑ [r/MachineLearning](#) · Posted by u/paul_read_it 2 days ago
- 121
↓
- [R] Tackling Climate Change with Machine Learning - video & blog post summary**
- Research
- ↑ [ironmagnesiumzinc](#) 8 points · 2 days ago
- ↓ This is fantastic. I'll definitely have to take a deep dive into this.
- Reply Give Award Share Report Save
- ↑ [jkff](#) 3 points · 1 day ago
- ↓ Thank you from the bottom of my heart. I already work on ML stuff, but ask myself every day how I can make my work potentially useful for saving the planet or whether it's time to go do something that's more directly useful for that (and what that would be - if I knew for sure, I likely would have already been knocking on their door). I've seen this paper before, but lacked the patience to read through it. It's wonderful to see it being put in a more digestible form.
- Reply Give Award Share Report Save
- ↑ [paul_read_it](#) 1 point · 17 hours ago
- ↓ Thank you for your kind words! Looking forward to find out what you'll be working on!
- Reply Give Award Share Report Save
- ↑ [bohreffect](#) 2 points · 2 days ago
- ↓ Thanks for posting this. Wrapping up my dissertation covering items 1-3 lumping it together as "civil infrastructure".
- Will definitely share these links with my classically trained engineering friends, as ML concepts and the increased availability of municipal/civil data is only now just starting to percolate out to traditional/physical PE type work.
- Reply Give Award Share Report Save
- ↑ [LavishManatee](#) 1 point · 2 days ago
- ↓ Thank you for posting this, I am fascinated.
- Reply Give Award Share Report Save

Aren't these comments beautiful? :)

Value-based feeds

Examples

Reddit (Machine Learning community)

↑ Posted by u/ladycrab 7 months ago
105 [R] A Generalization Theory of Gradient Descent for Learning Over-parameterized Deep ReLU Networks

↓ arxiv.org/abs/19...

Research

↑ chiasmodon 12 points · 7 months ago

↓ The paper is 50 pages, so I might have missed something. But from the highlighting in the text, Theorem 1.1 resp. 4.4 is the central result. IMO, there are three aspects of the claim:

- (i) By making the NN big enough, there exist parameter sets which achieve very low population loss. Qualitatively, this is ensured by the universal approximation property, and they impose something stricter about the ground truth (Assumption 4.2)
- (ii) A parameter set with low population error can be found based on a finite sample n of size as specified in the Theorem.
- (iii) Gradient descent works just fine in this regime, meaning that the non-convexity of the empirical loss surface is not a deal-breaker.

I think there has been a series of paper with results similar to (iii) recently, relying on the overparametrization $m \gg n$. Putting this in context with the generalization aspect (ii) is nice.

However, the sample complexity implied by their result is rather bad, corresponding to a rate of convergence $n^{1/6}$, which is much slower than the standard rate $n^{1/2}$. Also, there is no quantitative evaluation of their generalization bound. Hence, it can be hardly called a "non-vacuous generalization bound", in contrast to the cited work of Dzuigaite & Roi which yields an upper bound on MNIST of ~18% IIRC (Disclaimer: Im not an author). Also, "explaining the good generalization performance" might be overselling the result regarding the sample complexity.

Another comment: in Remark 4.7, the authors state that the "over-parametrization condition is independent ... of n ", yet they require $n > m^2$.

Despite the slow rate of convergence, the result claims that the classification problem is "learnable" by gradient descent on NN models. I am not qualified to assess the novelty of this claim, but it seems to be a satisfactory qualitative result.

Give Award Share Report Save

Paper summary

Review

People help each other in reviewing papers :)

Structure + value-based feeds

Examples

Should There be a Universal Basic Income (UBI)?

Wealthy countries should provide citizens with a universal basic income (UBI).

A UBI is more efficient than traditional welfare programs.

Pros

- A UBI helps reduce government bureaucratic and regulatory burdens.
- Employers will no longer be forced to play the role of social security guarantors. Therefore, employing people will be [easier, cheaper](#) and therefore more accessible.
- The reliability of a UBI, relative to other forms of welfare, allows the poor the ability to better plan and budget their spending.

Cons

- A blanket dividend such as UBI struggles to account for geographical disparities in the cost of living. In some regions, recipients might receive proportionally too much or too little as compared to other recipients in other areas.
- Traditional welfare programs allow governments to target specific issues in specific communities, a benefit that is not available with a UBI.
- Current welfare systems are already highly efficient. Empirical evidence shows that at [least 90%, if not 99%, of the capital](#)

Kialo

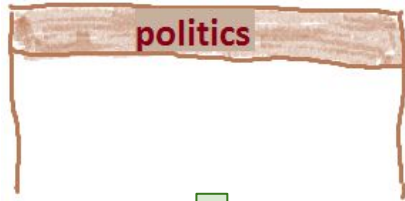
A total of 2700 arguments around the UBI question...

... All hierarchically structured and evaluated by peers.

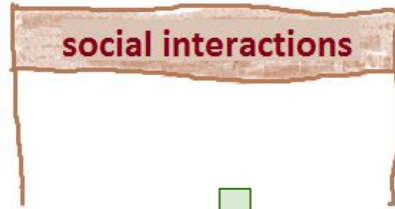
How can we apply this to open-source?

- 1) Transform project management in a collective task
→ **all active contributors** share power to merge code through voting
- 2) Improve cooperation around this (intellectual) task

How can collaborative thinking help open-source?



Easier group decision making
Transparent deliberation



Inappropriate language can
be downvoted



Management task can be
solved collaboratively

What's missing?

numpy / numpy

Issues 1,708

Want to contribute to numpy/numpy?

If you have a bug or an idea, read the contributing guidelines before opening an issue. If you're ready to tackle some open issues, we've collected some good first issues for you.

Filters: is:issue is:open

1,708 Open 5,971 Closed

- fftn very inefficient for 2d fft of several images
- Seq. fault with ufunc given NULL for 'data' in 'PyUFunc_FromFuncAndData', 00 - Bug, good first issue
- Minor typo in documentation for np.logical_and()
- piecewise is broken with x dtype=int64
- nanquantile does not return array of correct length if input is empty 00 - Bug, good first issue
- Update docs on extending numpy.random with Cython/Numba for 1.18.0 release 17 - Task
- Status of aliased/wrapped functions, including `numpy.cumproduct`, `np.alltrue`, `np.sometrue`, 07 - Deprecation
- issues with accuracy of quad precision? 57 - Closed
- DEP: consider Deprecating `arr.dtype = ...` attribute setting 07 - Deprecation 15 - Discussion
- Matrix multiplication much slower for ints than floats 01 - Enhancement 57 - Closed
- ImportError on Raspberry Pi 57 - Closed
- Numerical Error 57 - Closed

numpy / numpy

Pull requests 209

First time contributing to numpy/numpy?

If you know how to fix an issue, consider opening a pull request for it. You can read this repository's contributing guidelines or GitHub Help to learn how to open a good pull request.

Filters: is:pr is:open

209 Open 6,720 Closed

- DOC: Updated sphinx directive formatting
- MAINT: Bump pytest from 5.1.3 to 5.2.0 03 - Maintenance, component: CI
- MAINT: align random double_fill to float_fill, simplify distributions.c 03 - Maintenance
- WIP, API: rearrange the cython files in numpy.random 01 - Enhancement, 30 - API
- MAINT: Remove duplicate files from .gitignore
- BUG: properly define PyArray_DescrCheck(Exact) 00 - Bug, 09 - Backport-Candidate
- WIP, API: restructure and document numpy.random C-API 30 - API
- Use the fast histogram algorithm with provided bins, if possible
- DOC: Add to doc that interp cannot contain NaN
- DEP: Complete deprecation of invalid array/memory order
- MAINT: backport Cython API cleanup to 1.17.x, remove docs 03 - Maintenance
- MAINT: Propagate UnicodeDecodeError from PyArray_DescrConverter()
- DEP: Remove Python2 specific C module setup 27 - Py3K

What's missing? Value-based feeds

numpy / numpy

Issues 1,708

Want to contribute to numpy/numpy?

Filters: is:issue is:open

Issue Title	Labels	Comments
fftn very inefficient for 2d fft of several images		0
Seq. fault with ufunc given NULL for 'data' in 'PyUFunc_FromFuncAndData'	00 - Bug	1
Minor typo in documentation for np.logical_and()		0
piecewise is broken with x dtype=int64		4
nanquantile does not return array of correct length if input is empty	00 - Bug	2
Update docs on extending numpy.random with Cython/Numba for 1.18.0 release	17 - Task	0
Status of aliased/wrapped functions, including 'numpy.cumproduct', 'np.alltrue', 'np.sometrue'	07 - Deprecation	3
issues with accuracy of quad precision?	07 - Closed	8
DEP: consider Deprecating 'arr.dtype = ...' attribute setting	07 - Deprecation 15 - Discussion	7
Matrix multiplication much slower for ints than floats	01 - Enhancement 07 - Closed	3
ImportError on Raspberry Pi	07 - Closed	7
Numerical Error	07 - Closed	3

numpy / numpy

Pull requests 209

First time contributing to numpy/numpy?

Filters: is:pr is:open

Pull Request Title	Labels	Comments
DOC: Updated sphinx directive formatting		0
MAINT: Bump pytest from 5.1.3 to 5.2.0	03 - Maintenance component: CI	2
MAINT: align random double_fill to float_fill, simplify distributions.c	03 - Maintenance	4
WIP, API: rearrange the cython files in numpy.random	01 - Enhancement 30 - API component: numpy.random	11
MAINT: Remove duplicate files from .gitignore		2
BUG: properly define PyArray_DescrCheck(Exact)	00 - Bug 09 - Backport-Candidate	3
WIP, API: restructure and document numpy.random C-API	30 - API	2
Use the fast histogram algorithm with provided bins, if possible	Component: NumPy	4
DOC: Add to doc that interp cannot contain NaN		1
DEP: Complete deprecation of invalid array/memory order		11
MAINT: backport Cython API cleanup to 1.17.x, remove docs	03 - Maintenance component: numpy.random	1
MAINT: Propagate UnicodeDecodeError from PyArray_DescrConverter()		0
DEP: Remove Python2 specific C module setup	27 - Py3K	4

1. Understand why open-source is losing



2. Explore how we can rebalance forces

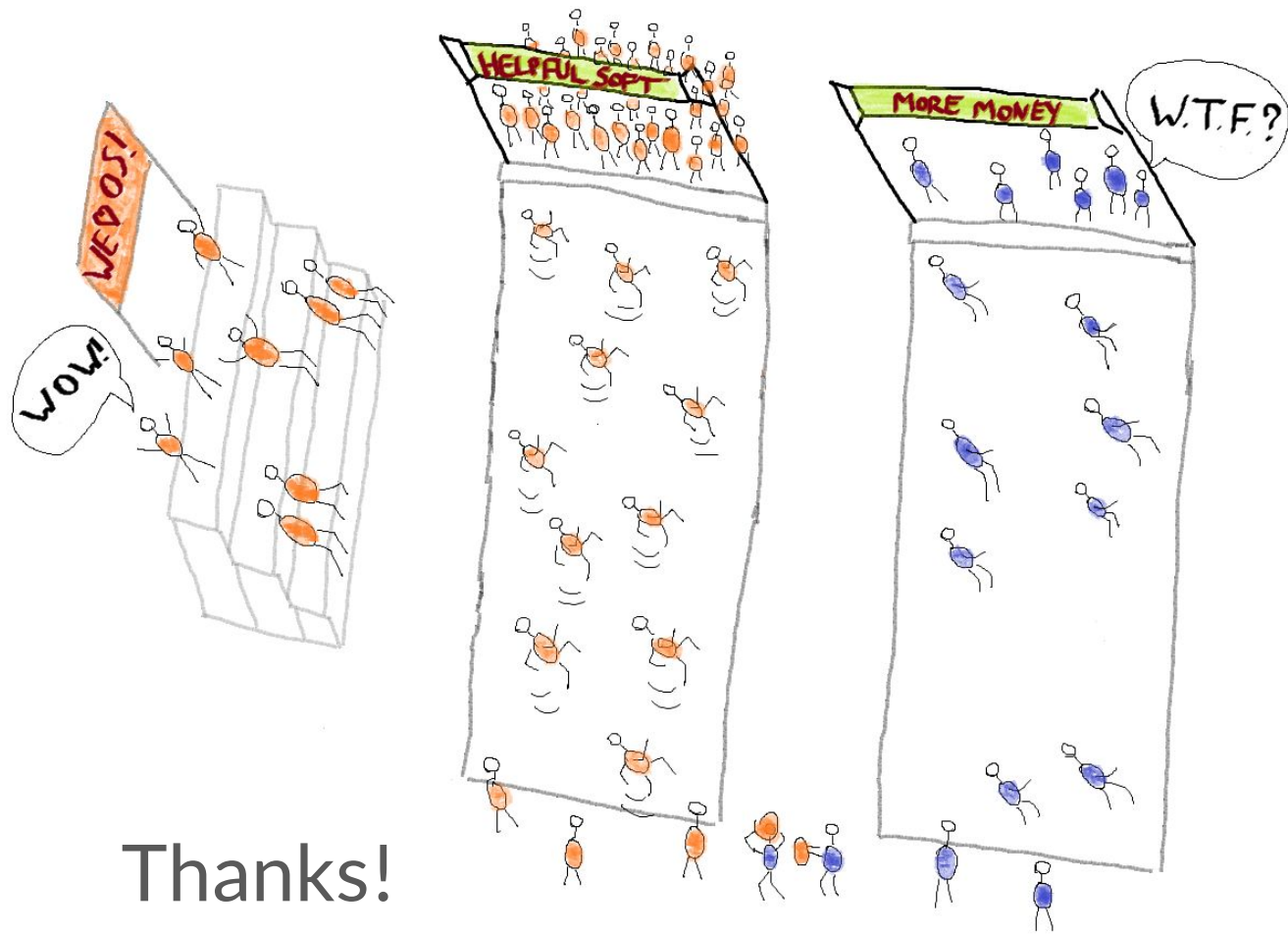


1. Understand why open-source is losing



2. Explore how we can rebalance forces





Thanks!

Thanks to...

Antoine, Arnould, Sam, Victor, Antoine,
Maxime, Anne-Sophie, Ploum,
my ex-flatmates (Mailingpot),
my flatmates (Cul-de-Sac), my family

...for the helpful discussions!

References for further readings....



definition open-source software: https://en.wikipedia.org/wiki/Open-source_software

definition proprietary software: https://en.wikipedia.org/wiki/Proprietary_software

some logo's: <https://www.flaticon.com/packs/software-development-logos>

Problems of open-source

libav- ffmpeg: https://www.theregister.co.uk/2015/08/05/ffmpeg_leader_steps_down/

<http://blog.pkh.me/p/13-the-ffmpeg-libav-situation.html>

neovim: <https://geoff.greer.fm/2015/01/15/why-neovim-is-better-than-vim/>

diaspora: <https://discourse.diasporafoundation.org/t/project-management-at-diaspora-or-what-the-heck-am-i-doing-here/563/8>

node.js: <https://www.zdnet.com/article/after-governance-breakdown-node-js-leaders-fight-for-its-survival/>

google: <https://fortune.com/longform/inside-googles-civil-war/>

facebook: <https://www.nytimes.com/2018/08/28/technology/inside-facebook-employees-political-bias.html>

diaspora: <https://discourse.diasporafoundation.org/t/the-core-team-a-bottleneck-of-the-diaspora-development/588>

felix kraus: <https://krausefx.com/blog/scaling-open-source-communities>

ffmpeg: <https://ffmpeg.org/pipermail/ffmpeg-devel/2015-July/176489.html>

burnouts: <https://techcrunch.com/2018/06/23/open-source-sustainability/>

roads and bridges:

<https://www.fordfoundation.org/about/library/reports-and-studies/roads-and-bridges-the-unseen-labor-behind-our-digital-infrastructure>

Solutions

open source governance <https://randyfay.com/content/how-do-open-source-communities-govern-themselves>

patreon: <https://www.patreon.com/>

opencollective: <https://opencollective.com/>

Definition DAO: https://en.wikipedia.org/wiki/Decentralized_autonomous_organization

aragon: <https://aragon.org/discover>

colony: <https://colony.io/>

daostack: <https://daostack.io/>

democracy.earth: <https://democracy.earth/>

Definition horizontal organization: https://en.wikipedia.org/wiki/Flat_organization

Reinventing organizations: <https://www.reinventingorganizations.com/>

Udn: <http://universite-du-nous.org/>

framadate: <https://framadate.org/>

trello: <https://trello.com/>

stackoverflow: <https://stackoverflow.com/>

reddit: <https://www.reddit.com/>

kialo: <https://www.kialo.com/>

Numpy's github: <https://github.com/numpy/numpy>

Value-based feeds

Examples

Stackoverflow

What's this?

55 Answers !

active oldest votes

1 2 next

9082

```
// Checks css for display:[none|block], ignores visibility:[true|false]
$(element).is(":visible");


// The same works with hidden
$(element).is(":hidden");
```


Same as [twernt's suggestion](#), but applied to a single element; and it [matches the algorithm recommended in the jQuery FAQ](#)

share improve this answer

edited Jul 2 '18 at 15:50

answered Oct 7 '08 at 13:30

 Black 5,525 ● 14 ● 68 ● 139

 Tsvetomir Tsonev 94.3k ● 3 ● 21 ● 32

