

SquirrelInHell

2016-10-23

Internal Race Conditions

第一

You might be familiar with the concept of a 'bug', as introduced by CFAR. By using the computer programming analogy, it frames any problem you might have in your life as something fixable... even more - as something *to be fixed*, something such that fixing it or thinking about how to fix it is the *first thing that comes to mind* when you see such a problem, or 'bug'.

Let's try another analogy in the same style, with something called 'race conditions' in programming. A race condition is a particular type of bug, that is typically very hard to find and fix ('debug'). It occurs when two or more parts of the same program 'race' to access some data, resource, decision point etc., in a way that is not controlled by any organised principle.

For example, imagine that you have a document open in an editor program. You make some changes, you give a command to save the file. While this operation is in progress, you drag and drop the same file in a file manager, moving to another hard drive. In this case, depending on timing, on the details of the programs, and on the operating system that you are using, you might get different results. The old version of the file might be moved to the new location, while the new one is saved in the old location. Or the file might get saved first, and then moved. Or the saving operation will end in an error, or in a truncated or otherwise malformed file on the disk.

If you know enough details about the situation, you could in fact work out what exactly would happen. But the margin of error in your own handling of the software is so big, that you cannot in practice do this (e.g. you'd need to know the exact millisecond when you press buttons etc.). So in practice, the outcome is random, depending on how the events play out on a scale smaller than you can directly control (e.g. minute differences in timing, strength of reactions etc.).

第二

What is the analogy in humans? One of the places in which when you look hard, you'll see this pattern *a lot* is the relation of emotions and conscious decision making.

E.g., a classic failure mode is a "commitment to emotions", which goes like this:

- I promise to love you forever
- however if I commit to this, I will have doubts and less freedom, which will generate negative emotions
- so I'll attempt to fall in love faster than my doubts grow
- let's do this anyway, why won't we?

The problem here is a typical emotional "race condition": there is a lot of variability in the outcome, depending on how events play out. There could be a "butterfly effect", in which e.g. a single weekend trip together could determine the fate of the relationship, by creating a swing up or down, which would give one side of emotions a head start in the race.

第三

Another typical example is making a decision about continuing a relationship:

- when I spend time with you, I like you more

Blog Archive

[January 2018](#) (3)
[December 2017](#) (3)
[November 2017](#) (1)
[October 2017](#) (2)
[September 2017](#) (1)
[August 2017](#) (2)
[July 2017](#) (1)
[May 2017](#) (2)
[April 2017](#) (1)
[March 2017](#) (2)
[January 2017](#) (2)
[November 2016](#) (1)
[October 2016](#) (1)
[September 2016](#) (2)
[August 2016](#) (1)
[April 2016](#) (1)
[March 2016](#) (1)

More by SquirrelInHell

- [AI Safety Comics](#)
- [Android Apps](#)
- [Be Well Tuned](#)
- [Rationality Updates](#)

- when I like you more, I want to continue our relationship
- when we have a relationship, I spend more time with you

As you can see, there is a loop in decision process. This cannot possibly end well.

A wild emotional roller-coaster is probably around the *least* bad outcome of this setup.

第四

So how do you fix race conditions?

By creating structure.

By following principles which compute the result explicitly, without unwanted chaotic behavior.

By removing loops from decision graphs.

First and foremost, by recognizing that leaving a decision to a race condition is *strictly* worse than any decision process that we consciously design, even if this process is flipping the coin (at least you know the odds!).

Example: deciding to continue the relationship.

Proposed solution (arrow represent influence):

(1) controlled, long-distance emotional evaluation -> (2) systemic decision -> (3) day-to-day emotions

The idea is to remove the loop by organizing emotions into two groups: those that are directly influenced by the decision or its consequences (3), and more distant "evaluation" emotions (1). A possibility to feel emotions as in (1) can be created by pre-deciding a time to have some time alone and judge the situation from more distance, e.g. "after 6 months of this relationship I will go for a 2 week vacation to my aunt in France, and think about it in a clear-headed way, making sure I consider emotions about the general picture, not day-to-day things like physical affection etc."

Total time to write this post: 31 minutes (23 wpm, 133 cpm)

No comments:

Post a Comment

[Newer Post](#)

[Home](#)

[Older Post](#)

Subscribe to: [Post Comments \(Atom\)](#)