

Kelly 2004: The Sunk Cost Fallacy

Kevin Dorst

24.223 Rationality

I. Honoring sunk costs

You're inclined to **honor sunk costs** \approx you take the fact that you (or others) have invested significant resources into achieving a goal as a further reason to do it.

The question: is this always (or usually) irrational?

The "sunk cost fallacy".

Examples:

- Read a book vs. go to theater. (\$200, non-refundable ticket.)
- Vietnam war. Can we simply dismiss this as the sunk cost fallacy, or must we argue against it more substantively?

The conventional wisdom:

- 1) **Descriptive:** People often do give weight to sunk costs in making their decisions; and
- 2) **Normative:** Doing so is irrational.

To preview Kelly's conclusion: What exactly counts as "giving weight to" sunk costs?

What if you do it to avoid regret?
What if to avoid perceptions of waste?

Certain reasons (like these) can clearly be rational to take account for, so if honoring sunk costs is to be irrational, must bracket them.

But then: how well can we determine what's *actually* motivating people, and know it's not the rational thing?

Most experiments don't control for this.

II. Why think it's irrational?

The past is the past! The factors that should determine your decision are the things you can *change*. You can't change the past, so it shouldn't affect your decision.

Idea: if had a memory-wipe pill that erased all and only knowledge about how many resources you've put into a project, would make better decisions without it.

Counterexamples:

- 1) *Evidential effects*. Past investments can be a good indicator of whether it will (or won't!) succeed.
- 2) *Outcome effects*. Past investments can affect things of (dis)value in the resulting outcome.

Should you try to finish cooking before class? Depends on how much you've done already.

E.g. will you feel regret? Will others think you are wasteful or fickle? Do you have a reputation to maintain?

Kelly: do (1) and (2) *exhaust* the rational effects that past investments can have on your decision-making? No.

Q: Can we understand what follows as simply delineating *another kind* of outcome effect? If EUT is right, then yes.

III. How to change the past

Actions now and their future consequences *can* change the (proper interpretation of) the past. And that can matter!

- Doctor working to save a stabbing victim determines whether the attacker is an assaulter or a murderer.
- Historical events often only achieve their significance much later.
- Within a life, later events can alter the whole narrative structure.

Kelly: as long as its rational to say a murder is worse than an assault, or the death of a potential-dictator is not as bad as the death of an potential-liberator, etc., then it can be rational to have preferences about the proper interpretation of past events.

So facts about sunk costs *can* rationally affect decisions.

Generally: often seems reasonable to have **redemptive preferences**, where we prefer that past sacrifices not be in vain.

E.g. your friend passes away, and you discover a manuscript she was working on secretly. Do you try to get it published? Surely how much she invested in it could affect your decision!

Q1: Is Kelly's reading of the above examples right?

Q2: Is it right that there's no first-/third-person asymmetry?

IV. When and why?

Suppose we agree that sometimes past sacrifices can affect the rationality of future decisions. If so, when and why?

When:

Kelly thinks this can only be rational in cases where there was some independent reason to achieve the goal in question.

Q: How to justify this?

- Bill cares about following through on his plans (prefers his past sacrifices not be in vain), *regardless* of whether the goal ends up having any independent value.
- When a goal has independent value, Jill cares about following through on her plans *more* than the independent value of the goal (because she prefers her past sacrifices *for valuable goals* not be in vain).

Seems like Kelly says that Bill is irrational but Jill is rational. Why?

He says that the rational tendency to honor sunk costs is driven by a respect for persons and their projects.

→ If we care about people and their projects, why should we only care about their *independently valuable* projects? *Do* we?

Survival of a young boy who grows up to be a dictator. Death of a patient that leads to a new type of treatment.

E.g. only reason you met your future partner was because you failed your exam.

And if this can be rational in third-personal cases, shouldn't it also be rational in first-personal ones?

E.g. searching for a book example.

V. The Upshot

Recall the conventional wisdom:

- 1) [Descriptive:] People often do give weight to sunk costs in making their decisions; and
- 2) [Normative:] Doing so is irrational.

Kelly says we can interpret “given weight to sunk costs” in either a *wide* or *narrow* way.

- Wide = any causal tendency for past investments to increase commitment to pursuing a goal.
- Narrow = the *mere fact* that there were past investments (regardless of evidential effects, outcome effects, or redemptive preferences) leads people to increase commitment to the goal.

On the wide interpretation, there is plenty of evidence for (1), but (2) is very often false.

Or, if true, not so *merely* because of (1).

On the narrow interpretation, (2) is true, but there is little if any evidence for (1).

Kelly suggests, even stronger: how could we get evidence for (1) on the narrow interpretation? Isn't it almost always more plausible/charitable to reinterpret “My reason is the past investment” as “my reason is the past investment, because I care about them not being wasted”.

Q: Is that right? Seems like it depends on how strong constraints we impose on what independently valuable projects are.

References

Kelly, Thomas, 2004. ‘Sunk costs, rationality, and acting for the sake of the past’. *Nous*, 38(1):60–85.