Values, Preferences, and Meaning[*]

Joe Edelman[†]

9 June 2021 — DRAFT

Abstract

Many fields (social choice, welfare economics, recommender systems) assume people express what benefits them via their 'revealed preferences.' Revealed preferences have well-documented problems when used this way, but remain fundamental to these fields because, as an information source, they are a *private*, *local*, *verifiable*, *fine-grained*, *battle-tested*, and *easily-expressed*. Alternative sources (about participants' values, capabilities and functionings, etc) can't compete. Or can they?

Introduction

The customer's always right. The people have spoken. These phrases show the central role of "revealed preference" in our markets and democracies. Policymakers work to serve citizens, driven by preferences revealed by ballots; product designers work to serve users-customers, driven by what's revealed by purchases and clicks.

Revealed preferences are called "engagement metrics" when they're summed up. A product or policy with high engagement is said to serve user-citizens most, because they chose it most. The people have spoken. The customer's always right.

This is alluringly simple. And yet: what we imagine with "values-based design" isn't exactly "engagement-maximizing design." Likewise, "values-based policy" isn't exactly populism.

Point this out, and people yell at you: Don't we know best for our own lives? Are you going to impose "better values" from above?

Here's a third option: What if people know best, and have wise values, but their engagements aren't the last word on those values? To prove this, you'd need another source of information on people's values. One that *is* the last word. But: what could be as robust as people's own revealed preferences?

Supposing we can figure *that* out, what do we mean by "values?"

Social visions. Sometimes people mean *visions of what's right for everyone, or for a group*—what a family should be like, how a father should behave, what a nation should be like, etc. On this definition, values would include things like inclusiveness, freedom, feminine dress-codes, etc.

Meaning nuggets. Other times, people mean *things that feel right and meaningful when you do them*—such as being vulnerable, taking stage, being creative, etc.

I believe we can get data about people's "meaning nuggets." As robustly as we get revealed preferences. We can use it to evolve designs and policies in service of values. To show this, I'll build on work by various philosophers and psychologists,[1] and say things about the role of attention in choice, meaningful experiences, and how theoretical knowledge differs from life wisdom. Because values touch on all of those things.

On Preferences

To evolve a product to suit users, a designer needs information about users—information that serves as evidence that a product direction is an improvement. Similarly, a policymaker needs information

---

[†]School for Social Design joe.edelman@gmail.com

[1]Sen, "Behaviour and the Concept of Preference"; Sen, "Rational Fools"; Taylor, "4 What Is Human Agency?"; Anderson, *Value in Ethics and Economics*; Chang, "'All Things Considered'"; Gibson, *The Senses Considered as Perceptual Systems*; Velleman, *Practical Reflection*.

1

about citizens, to justify policy directions.

Various information is collected, but the gold standard is revealed preferences, because they are:

- **Verifiable**. Engagements leave a trail. Who did you actually vote for? What did you purchase?
- **Local**. Preferences are informed by the local situations and priorities of user-citizens. Most alternatives assume outside experts know better than user-citizens what would help them.
- **Battle-tested**. We often *say* we want things, but don't choose them in the final analysis. Preferences get at our *real* priorities.
- **Fine-grained**. Preferences say subtle things about how a person wants to live.
- **Private**. Preferences are often expressed away from the signaling of allegiances, so they're less influenced by social pressure.
- **Easily-expressed**. You don't need to be particularly introspective or good with words to act on a gut feeling about what's right for you.

Despite these advantages, we have terms like *addiction, soulless consumerism, atomization*, and *populism* that describe how revealed preferences, summed up into engagement metrics, lead us astray. The problem is this: revealed preferences omit signals we'd rather collect, and collect noise we'd rather set aside:

Lost signal. Alex wants to move to a different city, but only if his friends also move.[2] Alice and her friends have, unfortunately, revealed a preference for their *current* city. Ben and *his* friends want to play tennis more—but they're choosing individually, from a menu of bookable tennis experiences. It looks like a preference to play tennis separately—even a *rivalry*[3] for the same tennis court.

Noise. Carla buys a car because there's no local transit. This counts as a preference for cars. Dan does something he later regrets, due to social pressure, misinformation, or a manipulative ad. Preference! Eli is following a social rule which she knows is *not* in her best interest.[4] Preference!

In general, choices made out of thoughtlessness, misinformation, lack of coordination, or external pressure count the same as those from reflection, experience, capacity, and wisdom.

This all stings worse in some parts of life: those concerning *community* and *meaning*. That's why engagement-optimizing systems—like markets, democracies, and social media—aren't where we turn for meaning and community.

The Nature of Choice

It's worth looking deeper at that lost signal and noise. It happens when the options we have in mind to choose from are biased or incomplete. To solve this, we can widen our conception of choice to include option-set formation.[5]

Imagine I'm with colleagues, and I say something witty. At t_1, I choose to say *that* in particular. But, at some earlier point t_0, I decided to put some attention on finding witty things to say. Since then, I try witty ideas in my mind, and consider situations around me as opportunities for wittiness. My attentional policy (adopted at t_0) of 'looking for witty quips and reframes' is how I assembled the option set for the latter choice, at t_1, when I choose from the best I've found.

At t_0, I adopted an **attentional policy**. To get more specific about this: by **policy**, I mean something like "taking out the trash when it's almost full," "calling mom on Sundays," or "running new contracts past the lawyer"—something done regularly, or in a certain context, without a cost-benefit analysis each time.[6]

Attentional policies (APs), then, are policies about how to *think* about a thing, what to *pay attention to* in a context, or *what to look for in selecting an action*. "Taking out the trash on Tuesdays" is a normal policy, but "experiencing every step and breath while doing my chores" or "look-

---

[2]See discussions of the prisoner's dilemma in e.g., Sen, "Behaviour and the Concept of Preference"; Anderson, "Symposium on Amartya Sen's Philosophy."

[3]Edelman, "Towards 'Game B'."

[4]See Sen, "Rational Fools"; Anderson, *Value in Ethics and Economics* on 'commitment.'

[5]See Smaldino and Richerson, "The Origins of Options" for a great review.

[6]Policies are necessary (just as plans, heuristics, norms, intentions, etc) due to bounded rationality and social coordination. (Bratman, *Intention, Plans, and Practical Reason*; Simon, "A Behavioral Model of Rational Choice"; Taylor, *The Language Animal*)

ing for kind words when giving feedback"—*these* are attentional policies.

APs can be about how to treat people[7] (*honestly, openly, generously, mercilessly*); how to approach things (*with reverence, with levity, with skepticism*); how to keep things (*simple, sensual, rocking, full of surprise*); or how to act more generally (*boldly, thoughtfully, carefully*).

TREAT PEOPLE	KEEP THINGS	ACT	APPROACH THINGS
honestly	simple — like Beck, like Dieter Rams	boldly — like Indiana Jones, like Rosa Parks	with reverence — like Michelangelo
openly	sensual	thoughtfully — like Moriarty, like Obama	with transparency
generously — like Al Capone, like Oprah, like Mother Teresa	rocking	carefully	as a performance
without mercy	full of surprise	wisely	with levity — like Kanye West, like Jim Carey, like Ellen
with compassion	free	forcefully — like Darth Vader, like Andre Agassi	with faith
with loyalty	equal	with calculated mystery	with curiosity — like Sherlock Holmes, like Oscar Wilde
with devotion	fair — like kindergarten	with precise timing — like Bach, like George Soros	
	rational		
	just		

Figure 1: Some attentional policies

For replacing preferences, APs look good:

1. **Fine-grained.** To guide our attention, APs must be precise. "Be honest" is too vague—it doesn't tell me what to look for. So, a policy like "be honest" is short for a more specific articulation, like "attend to what I feel about each thing we discuss, and let my feeling show," or "attend to any false impressions the listener might get from my statements, and head them off with a disclaimer." To have honesty as a policy, I must first have a substantive interpretation of honesty.

2. **Local.** These substantive interpretations differ from person to person. In fact, APs make up much of what we call "personality": When making friends, are you cautious or bold? When considering a purchase, is the focus on price, quality, or durability? When speaking, do you try to be witty, precise, or down to earth? Often these aren't just "character traits" a person is born with, but policies adopted for a reason, which work together for that person's way of life.[8]

3. **Hard-boiled.** If I had a galaxy brain, I'd have a million attentional policies, all in the same context. Talking to colleagues at work, I might craft my words to be kind, honest, tactful, humble, and inspiring—and try to be precise in my speech, aware of how each word lands, aware of my own feelings, and transparent with them. Calm and centered, but also passionate. Physically graceful, like a dancer.

This is impossible, since policies *compete* for my attention. Instead, I must choose (often intuitively, unconsciously) what to attend to, in each context.[9]

A decision to look for witty things to say means I'm *not* looking for vulnerable things to say, or helpful things, or mysterious things, etc. That's a choice *not* to be vulnerable, helpful, or mysterious!

These are some of the biggest choices we make. The opportunity costs are so high that many policies which *sound* very good (like being completely present, or endlessly compassionate) don't win out in the real contexts of life.

4. **Verifiable.** If someone says they have an AP *a* in context *c*, you can put them in *c* and see what they attend to, or relatedly, what options they find. Alternatively, you can test for detailed knowledge of when exactly it makes sense for them to follow *a*. If they really have *a*, they'll know the shape of *c* well.

Justification Shapes

Some APs seem like "values"[10]. But not all. For instance, this one doesn't:

> I'm careful with my speech at work. My boss fires anyone who speaks imprecisely.

But this one does:

> I recently opened up to a friend about a struggle of mine. Since then, the relationship feels more intimate, and stronger; it's

[7]APs may account for the rationality of collective action better than the tricks in Anderson, "Symposium on Amartya Sen's Philosophy"; This approach to social choice and the P.D. follows Velleman, *Self to Self*, ch 11, closely.

[8]The characterization of identity as composed of APs shares something with **sAnderson2001?**; and Velleman, *Practical Reflection*.

[9]I have an idea *when* to try for wittiness: maybe at work, but not in a fight with my wife.

[10]Values are often considered as evaluative criteria or attitudes (Chang, "'All Things Considered'", Velleman, *Practical Reflection*). Here, I treat them as policies, but I think these definitions are interchangeable. An evaluative attitude or criterion can be viewed as something a person does when making an evaluation or choice.

easier for me to think about what to say; my friend is unexpectedly helpful. Now, I can't imagine a good life that doesn't include being honest with friends.

One difference is how they're justified: speaking carefully is justified by a chain of hypotheses about the consequences. I must speak carefully, or I'll say things my boss doesn't like; then, I'll get fired. We can visualize these reasons as a chain X⟶Y⟶Z.

In the second story, my honesty leads to all sorts of benefits, not chained together. We can visualize these reasons (which point in many directions) as a star ⋆.

When reasons for a policy form a chain ⟶, I'll say it's *narrow-justified*, or adopted for *narrow-benefits*. When the reasons are a star ⋆, I'll say *broad-justified* or *broad-benefits*.

I'll only call something broad-justified if it's adopted for *bountiful*, *redundant*, and *untracked* reasons. By *bountiful,* I mean I haven't listed them all. I've only *started* listing benefits of honesty and expect to discover new ones. By *redundant*, I mean I'd continue being honest if any one benefit (such as "my friend is unexpectedly helpful") turned out mistaken. By *untracked* I mean that, when being honest, I'm not tracking whether the benefits I've named happen in each case. I just focus on being honest.[11]
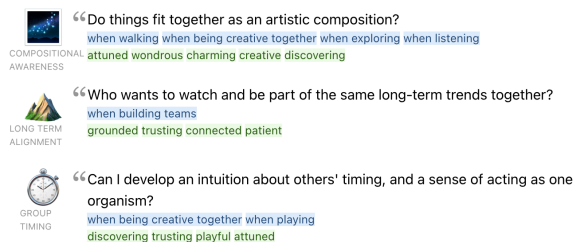
COMPOSITIONAL AWARENESS — "Do things fit together as an artistic composition? when walking when being creative together when exploring when listening attuned wondrous charming creative discovering

LONG TERM ALIGNMENT — "Who wants to watch and be part of the same long-term trends together? when building teams grounded trusting connected patient

GROUP TIMING — "Can I develop an intuition about others' timing, and a sense of acting as one organism? when being creative together when playing discovering trusting playful attuned

Figure 2: Examples of broad-justified APs, from our database, meaning.supplies.

In the intro, I mentioned two kinds of "values": social visions and meaning nuggets. Here's a social vision:

Andrew believes a pervasive dishonesty is undermining democracy and civil soci-ety. For this reason, he tries to spread honesty—denouncing lies, and setting an example of honesty wherever he goes.

Andrew thinks honesty will save democracy, and that he should spread it by being honest himself. This is a chain of hypotheses ⟶. In general, social visions create narrow-justified APs (NAPs), not broad-justified ones (BAPs).

If you want meaning nuggets, without social visions[12], select only BAPs. This will also filter out other things—what's done just to keep our jobs, fit in with a friend group, achieve specific goals, or get good sensations—what other philosophers call instrumental goods.[13]

Wisdom and Meaning

This ⟶ vs ⋆ thing is powerful. It can separate our knowledge into two piles. Knowledge towards narrow benefit is *know-how;* well-informed broad justifications are *wisdom*.

> ***Wisdom**, n.* Knowing from experience which policies are broadly beneficial.

Broad justifications are harder to communicate— they're made of many data points, usually collected via experience living a certain way. That's why "life wisdom" mostly comes from experience. No matter how dog-eared your Kahlil Gibran book is, you haven't collected all those diffuse benefits.

When wisdom *does* come quick, it's often via *meaningful experience.* Like this:

> $t = 1$. Brenda sips her morning tea, watching a bird on the feeder. Something shifts in her; she sees the bird shares a great project with her. She and the bird

---

[11] This argument descends from Boyd, "How to Be a Moral Realist."

[12] Social visions often feel meaningful, but only because they ride on a companion BAP. E.g., my social vision of *spreading values-articulacy* rides on a BAP: *responding deeply to the world-situation as I find it.* To verify that the meaning comes from the BAP, I can check for vison-but-not-BAP meaningless times, and vice versa.

[13] Or: what's broad-justified *constitutes* the good life; what's narrow-justified is done *to reach* the good life, from a position outside it. This makes some geometric sense, if reasons are steps in a topological space. In any case, I hope broad vs narrow is more defensible than intrinsic vs instrumental.

are explorers and representatives of what it is to be alive.

$t = 2$. Over time, thinking of herself as "an explorer of what it is to be alive" becomes a new kind of curiosity for Brenda. It comes up when she "does animal things" (in the woods, overcome by emotion, plunging into cold water). She notices more about her environment, and about how she feels.

$t = 3$. Brenda's become more explorative and bolder. Her attention shifts: when in nature, etc, she no longer focuses on being an explorer, but on balancing exploration with other factors, like self-care. Being "an explorer of what it is to be alive" is still, in a sense, something she does. It's still important to her. But it's become *automatic.*

At $t = 1$, Brenda might say her bird-moment is a new idea about something broadly-beneficial. But she doesn't see how to repeat it, or further explore it.

At $t = 2$, Brenda has a new BAP. She probably doesn't have a phrase for it, like "being an explorer of what it is to be alive." But she's (a) adopted a new mode of attention for some contexts, and (b) feels it has many benefits, (c) without tracking them, or (d) hold any one benefit necessary to justify the BAP.

At $t = 3$, "greeting the world as an explorer of what it is to be alive" has ceased to be meaningful. It may *become* meaningful again, if she loses her way—for instance, if she gets too busy with work, or loses touch with her curiosity.

So: at $t = 1$, it's meaningful. At $t = 2$, it's meaningful and wise. At $t = 3$, it's wise, but no longer meaningful. In other words, Brenda moves from left to right in this chart.

What's meaningful is the attentional and broadly-beneficial. Experiences of meaning get distilled into reasons to adopt BAPs. We call those reasons *wisdom.* Meaning is the frontier of wisdom—the part we still need attention for.
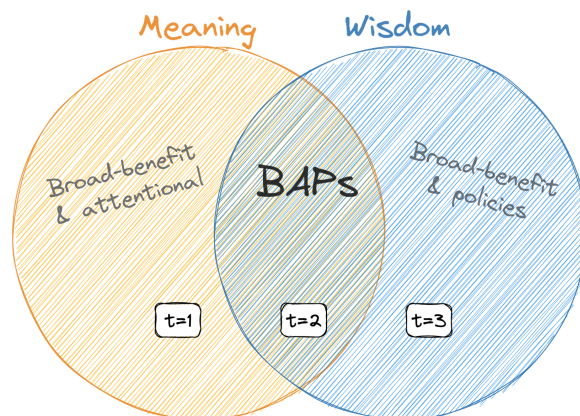
Or just:

Figure 3: Brenda's Journey

Meaning, the first derivative of a wise, good life.

Conclusion

"**The customer's always right.**" "**The people have spoken.**" *How about a new one?*

> Help others live by their own wisdom and sense of meaning.

Diverse problems stem from engagement-maximizing systems. Depression, media clickbait, isolation of the elderly, obesity, over-consumption, political polarization, bullshit jobs—all stem from a gap between preference and what people find meaningful and wise.

To close that gap, we can start by using BAPs as a way to collect what's meaningful to people, and their life wisdom. We do this at the School for Social Design[14]: we ask people what they attend to in various contexts[15], then remove the NAPs.

Then, instead of asking if users *engage* with a product, we ask if it makes space for them to be vulnerable, to be "explorers of what it is to be alive," or whatever their BAPs are.

Although this gets collected individually[16], BAPs

---

[14]https://human-systems.org/school

[15]E.g., what do they attend to while socializing? Is it how much space they give the people around them? Is it saying things carefully, so that their friends aren't upset?

[16]And thus leave some problems with preference-based

reduce problems of atomization. They capture what sort of social life individuals find meaningful—information often missing from preferences due to coordination problems. And BAPs contain wisdom, also often lost in preference, about the smooth operation of social life. For instance, a scientist's BAPs might include intellectual humility, passionate pursuit of the truth, etc. These aren't just meaningful for scientists, they keep the institution of science on the rails.

One next step is to explore BAPs empirically. Are they more or less stable than preferences? How often are they regretted? How independent are they from social pressure, framing effects, or the specifics of survey language?

Another next step is to make social choice mechanisms that augment preferences with BAP-information. For instance, by taxing BAP-negative outcomes, or subsidizing BAP-positive ones.

> Carla has a vulnerability-BAP, and changes workplaces. Her new "hanging out with colleagues" context is different from the previous one. How do they stack up for vulnerability? She says she experiences BAP-losses (without any corresponding BAP-gains) at her new job.

Does Carla's job-change mean her new job is better? Or is the rent rising? Was she misinformed? BAP-information can show this, where preference information wouldn't. Transactions which undermine BAPs, like Carla's job switch, could be discouraged (e.g., through taxation), or more meaningful workplaces could be subsidized.

Such mechanisms must be made robust[17], and may depend on streamlining BAP collection: BAPs are not so easily-expressed as preferences. Can interfaces for assisted introspection bring the articulacy requirements of BAPs closer to those of preferences? Can BAPs be inferred from other data?

Time will tell. For now, BAPs are leading to new and beautiful designs, and that's enough.

---

measures unsolved—such as agnosticism about distribution, and difficulty with interpersonal comparison.

[17]Laws and auditing systems exist to avoid preference fraud (collusion, cooked books, snake oil, voting fraud, etc). Can BAPs be similarly secured? In Edelman, "How to Operationalize Metaethics." I argue they can be *more* secure.

References

Anderson, Elizabeth. "Symposium on Amartya Sen's Philosophy: 2 Unstrapping the Straitjacket of 'Preference': A Comment on Amartya Sen's Contributions to Philosophy and Economics." *Economics and Philosophy* 17, no. 1 (2001): 21–38. https://doi.org/10.1017/S0266267101000128.

———. *Value in Ethics and Economics.* Harvard University Press, 1993.

Boyd, Richard. "How to Be a Moral Realist." In *Essays on Moral Realism*, edited by G. Sayre-McCord, 181–228. Cornell University Press, 1988.

Bratman, Michael. *Intention, Plans, and Practical Reason.* Cambridge, MA: Harvard University Press, 1987.

Chang, Ruth. "'All Things Considered'." *Philosophical Perspectives* 18, no. 1 (2004): 1–22. https://doi.org/10.1111/j.1520-8583.2004.00018.x.

Edelman, Joe. "How to Operationalize Metaethics." https://www.notion.so/humsys/How-to-Operationalize-Metaethics-2919bbd6a882487fb3083029feaceffa, 2020.

———. "Towards 'Game B'." https://www.notion.so/humsys/Towards-Game-B-dab1d8f5e0f14c7991939c07a0a3b488, 2020.

Gibson, James. *The Senses Considered as Perceptual Systems.* Westport, Conn: Greenwood Press, 1983.

Sen, Amartya. "Behaviour and the Concept of Preference." *Economica* 40, no. 159 (August 1973): 241. https://doi.org/10.2307/2552796.

Sen, Amartya K. "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory." *Philosophy and Public Affairs* 6, no. 4 (1977): 317–44.

Simon, Herbert A. "A Behavioral Model of Rational Choice." *The Quarterly Journal of Economics* 69, no. 1 (February 1955): 99. https://doi.org/10.2307/1884852.

Smaldino, Paul E., and Peter J. Richerson. "The Origins of Options." *Frontiers in Neuroscience* 6 (2012). https://doi.org/10.3389/fnins.2012.00050.

Taylor, Charles. "4 What Is Human Agency?" In *The Self: Psychological and Philosophical Issues*, edited by Theodore Mischel, 103. Rowman &

Littlefield, 1977.

———. *The Language Animal: The Full Shape of the Human Linguistic Capacity.* Harvard University Press, 2016.

Velleman, David. *Practical Reflection.* Princeton University Press, 1989.

Velleman, J. David. *Self to Self: Selected Essays.* Cambridge University Press, 2005.