Formal Value Theory

Kenny Easwaran
Texas A&M University

July 13, 2018

Main question — when is one thing better than another? How does overall betterness relate to the individual dimensions of betterness?

Examples:

- Multi-criteria decisions (dinner entrees can be compared on price, flavor, nutrition, and other criteria; careers can be compared on stability, pay, flexibility, interest, etc.)
- Elections (some voters think John Quincy Adams is better than Henry Clay is better than Andrew Jackson, while other voters have a different ordering)
- Decisions under uncertainty (if the market crashes, then buying gold is better than buying a mutual fund which is better than buying Amazon stock; if the market booms one way, then buying the mutual fund is better than buying Amazon which is better than gold; other possibilities yield other orderings)
- Social welfare and a theory of justice (high taxes and universal healthcare is very good for the poor and moderately good for the rich; low taxes and private healthcare is bad for the poor and very good for the rich; pure anarchy is bad for everyone)

Properties of a formal relation:

- Transitive: $\forall x \forall y \forall z((xRy \wedge yRz) \rightarrow xRz)$
- Reflexive: $\forall x(xRx)$
- Irreflexive: $\forall x \neg(xRx)$
- Symmetric: $\forall x \forall y(xRy \rightarrow yRx)$
- Anti-symmetric: $\forall x \forall y(xRy \rightarrow \neg yRx)$
- Total: $\forall x \forall y(xRy \vee yRx)$

A relation is called a "weak partial ordering" iff it is reflexive and transitive. A relation is called a "strict partial ordering" iff it is irreflexive and transitive. A relation is called a "linear ordering" iff it is a partial ordering and is total. A relation is called an "equivalence relation" iff it is reflexive, transitive, and symmetric.

If $\succeq$ is a weak partial ordering, define $x \approx y$ iff $x \succeq y$ and $y \succeq x$.

Theorem: $\approx$ is an equivalence relation.

We say R_2 "extends" R_1 iff $\forall x \forall y(xR_1y \rightarrow xR_2y)$.

Theorem: For every partial ordering, there is a linear ordering that extends it.

Goal: For a collection of partial orderings representing the different dimensions of value (criteria, voters, possibilities, individuals) find a partial ordering representing overall value.

1 Arrow's impossibility theorem

- Let f be a function that assigns a group linear ordering to a sequence of linear orderings for each voter.
- Say that f has "universal domain" iff it gives a unique linear ordering for *every* "profile" (an assignment of one linear ordering for each voter).
- Say that f obeys "independence of irrelevant alternatives" if the group ordering of A and B doesn't change unless some individual changes their ordering of A and B.
- Say that f obeys "unanimity" iff $A \succ B$ in the group ordering whenever $A \succ B$ in all individual voter orderings.

Theorem: Consider a set of 3 or more candidates, and a set of 3 or more voters. If f satisfies universal domain, unanimity, and independence of irrelevant alternatives, then there is some individual (the "dictator") such that the group ordering is always identical to that individual's ordering.

2 Numerical valuation

A "value function" assigns a real number to every element of the domain.

A value function V "represents" a linear ordering iff $\forall x \forall y (xRy \leftrightarrow V(x) > V(y))$.

Theorem: Any linear ordering on a finite set is represented by infinitely many distinct value functions. There exist linear orderings (on infinite sets) that aren't represented by any value function.

Say that f satisfies "anonymity" for two individual orderings $\succ_i$ and $\succ_j$ iff interchanging the two orderings always produces the same group ordering.

If we let f depend on the value functions rather than the orderings, then we can say that f satisfies "trade-off equality" for two individual value functions iff changing the value on those two functions by equal and opposite amounts results in an option that is overall equally good.

Theorem: If there are finitely many individual orderings, then the overall ordering satisfies trade-off equality for every pair of individual orderings iff it always agrees with the ordering based on the sum of values.