Todd Davies, **Decision Behavior: Theory and Evidence** (Spring 2010)
Version: April 7, 2010, 4:05 pm

3. Preference

3.1 Preference Relations

The development of preference theory below follows that of Kreps (1988).

DEFINITION 3.1.1. An *n-ary relation R* on sets $\boldsymbol{S_1,S_2,\ldots,S_n,}$ is a subset of the cross-product $\boldsymbol{S_1 \times S_2 \times \ldots \times S_n}$ (written as $R \subseteq \boldsymbol{S_1 \times S_2 \times \ldots \times S_n}$). The relation R holds between $s_1,s_2,\ldots,s_n$ (or $R(s_1,s_2,\ldots,s_n)$) iff the ordered pair $<s_1,s_2,\ldots,s_n> \in R$.

- Note: If R is a *binary* relation (n=2) that holds for $<s_1,s_2>$, we may also write s_1Rs_2.

Let $\boldsymbol{X}$ be any set of *possibilities* or *outcomes*. An outcome may be thought of as a proposition which specifies an event that may happen in the future. Outcomes should be comparable to each other, and may, but need not, be mutually exclusive events.

EXAMPLE 3.1.2. $\boldsymbol{X}$ = {Sally receives a piece of chocolate cake tomorrow at lunch, Barack Obama is elected President of the United States in 2012, A Republican is elected President of the United States in 2012}.

Let $P_i \subseteq \boldsymbol{X \times X}$, where $<x_1,x_2> \in P_i$ denotes that for agent *i,* outcome x_1 is strictly preferred to outcome x_2. When speaking generally, we may omit the reference to an agent, and write simply: x_1Px_2 (x_1 is strictly preferred to x_2). Let $x_1\neg Px_2$ denote that x_1 is *not* strictly preferred to x_2.

DEFINITION 3.1.3. $P \subseteq \boldsymbol{X \times X}$ is a (strict) *preference relation* iff
(a) $\forall x,y\ xPy \Rightarrow y\neg Px$ (*asymmetry*),
and
(b) $\forall x,y,z\ \ x\neg Py\ \&\ \ y\neg Pz \Rightarrow x\neg Pz$ (*negative transitivity*).

THEOREM 3.1.4. If $P \subseteq \boldsymbol{X \times X}$ is a preference relation then $\forall x,y,z\ \ xPy\ \&\ \ yPz \Rightarrow xPz$ (*transitivity*).
Proof (Halpern, 2006). Suppose that xPy, yPz, and $x\neg Pz$. By asymmetry, $z\neg Py$. But by negative transitivity, $x\neg Py$, contradicting our assumption.

EXPERIMENT 3.1.5. *Intransitive preferences.* People's preferences violate transitivity in some contexts. Subjects in Tversky (1969) were asked to express pairwise preferences between applicants for college admission. Applicants were described on three dimensions: intellectual ability, emotional stability, and social facility. Subjects selected one applicant from each pairing from a set of five applicants. There were four sets of profiles, shown below in a table reproduced from the original paper:

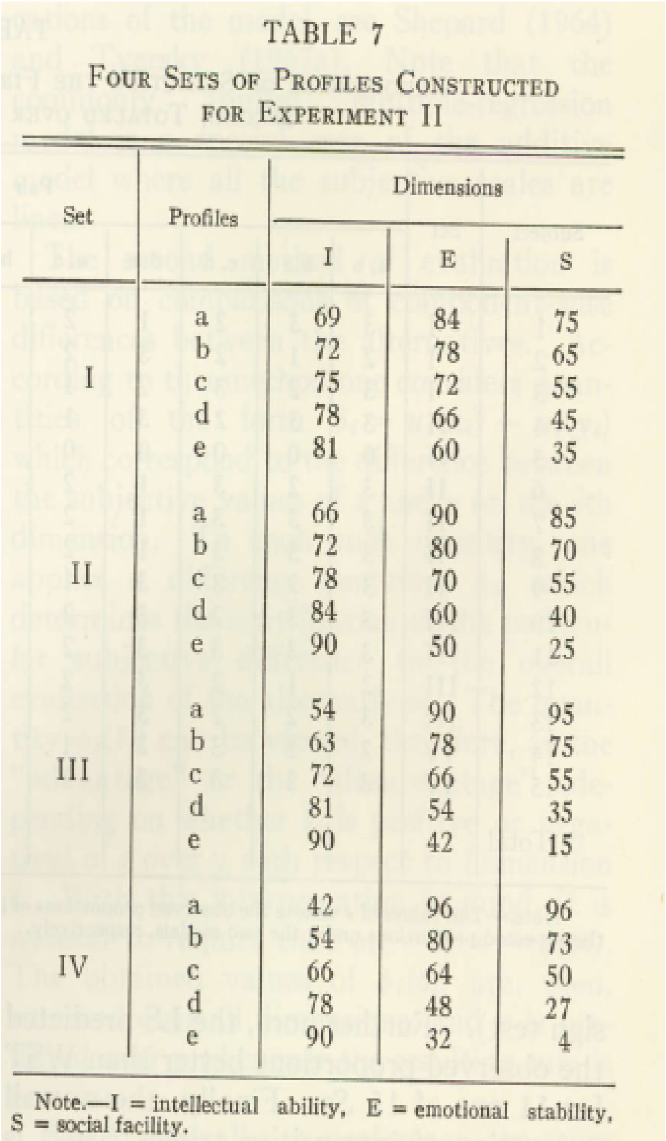


TABLE 7

FOUR SETS OF PROFILES CONSTRUCTED FOR EXPERIMENT II

Set	Profiles	Dimensions		
		I	E	S
I	a	69	84	75
	b	72	78	65
	c	75	72	55
	d	78	66	45
	e	81	60	35
II	a	66	90	85
	b	72	80	70
	c	78	70	55
	d	84	60	40
	e	90	50	25
III	a	54	90	95
	b	63	78	75
	c	72	66	55
	d	81	54	35
	e	90	42	15
IV	a	42	96	96
	b	54	80	73
	c	66	64	50
	d	78	48	27
	e	90	32	4

Note.—I = intellectual ability, E = emotional stability, S = social facility.

The design of the experiment was somewhat complicated, involving a preliminary and a test session, and some distractor profiles. Subjects expressed preferences three times for each pairing in the test phase. Responses were compared to models of choice probabilities predicted by adherence to weak stochastic transitivity (WST[1]) or another procedure (the lexicographic semiorder or "LS") that violates transitivity. Eleven out of 15 test subjects' responses were better modeled by the transitivity-violating model LS than by the transitivity-obeying model WST. When interviewed afterward, none of the subjects were aware that their preferences were intransitive. The better performing descriptive model (LS) assumes that subjects regard applicants as equivalent on a dimension if their scores on that dimension are within a threshold of difference, but select based on the most important dimension if the difference between applicants on that dimension is above the threshold. Dimension I was assumed to be the most important for this task, so that when differences were above threshold, subjects would be expected to weight that dimension highly in selecting between applicants. The four profile set groups in the test phase were assigned

[1] WST is defined as holding when it is the case that if the probabilities of selecting *x* over *y* and *y* over *z* are both above ½, then the probability of selecting *x* over *z* is also above ½.

based on data from the preliminary experiment in which each subject's threshold was estimated, and the profiles in each set were constructed so that the differences between alphabetically adjacent applicants in each set on dimension I were within the threshold estimated for that subject group.

EXPERIMENT 3.1.6. *Preference reversal – choice versus pricing*. Tversky, Slovic, and Kahneman (1990) had subjects both choose between pairs of options (e.g. a long-term prospect such as $2500 5 years from now or a short-term prospect such as $1600 1.5 years from now) and state “the smallest immediate cash payment for which they would be willing to exchange the delayed payment” (the cash equivalent *C*). The two procedures were done in opposite orders with different subjects, for options with long-term, short-term, and immediate payoffs. The results are shown below in a table reproduced from their paper:

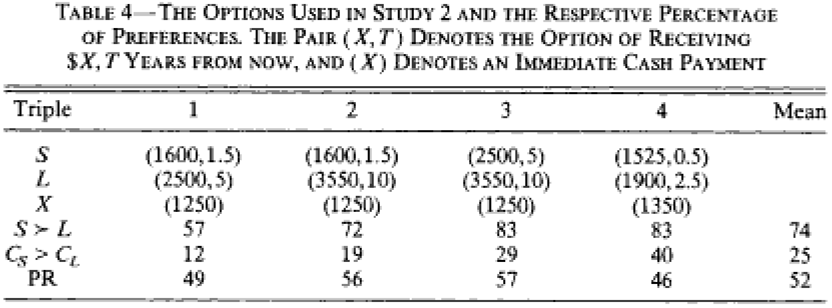
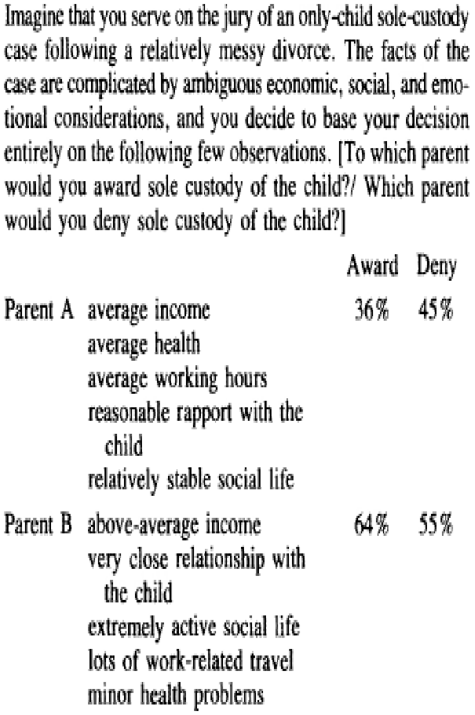


TABLE 4—THE OPTIONS USED IN STUDY 2 AND THE RESPECTIVE PERCENTAGE OF PREFERENCES. THE PAIR (X, T) DENOTES THE OPTION OF RECEIVING $$X, T$ YEARS FROM NOW, AND (X) DENOTES AN IMMEDIATE CASH PAYMENT

Triple	1	2	3	4	Mean
S	(1600,1.5)	(1600,1.5)	(2500,5)	(1525,0.5)	
L	(2500,5)	(3550,10)	(3550,10)	(1900,2.5)	
X	(1250)	(1250)	(1250)	(1350)	
$S \succ L$	57	72	83	83	74
$C_S > C_L$	12	19	29	40	25
PR	49	56	57	46	52

As Tversky et al. state: “Table 4 reveals a massive amount of PR [preference reversal]. Overall, the short-term option (*S*) was chosen over the long-term option (*L*) 74 percent of the time, but was priced higher only 25 percent of the time, yielding more than 50 percent PR patterns” (p. 213). This violates the asymmetry axiom of strict preference if we assume people prefer more money to less money.

EXPERIMENT 3.1.7 *Preference reversal – accept versus reject.* A series of studies done by Shafir (1993) illustrate that people's apparent preferences between two outcomes can change depending on whether they are being asked a positive or a negative question about the two options. In one problem (shown below), the majority answer switched from one parent to the other in a child custody decision, depending on whether subjects were asked which parent should be “awarded” or “denied” custody.



Imagine that you serve on the jury of an only-child sole-custody case following a relatively messy divorce. The facts of the case are complicated by ambiguous economic, social, and emotional considerations, and you decide to base your decision entirely on the following few observations. [To which parent would you award sole custody of the child?/ Which parent would you deny sole custody of the child?]

		Award	Deny
Parent A	average income average health average working hours reasonable rapport with the child relatively stable social life	36%	45%
Parent B	above-average income very close relationship with the child extremely active social life lots of work-related travel minor health problems	64%	55%

These and many other experiments done with chance lotteries show that people's apparent preferences between outcomes differ depending on the elicitation procedure (choosing versus pricing, or accepting versus rejecting). The proposed explanation is that people's preferences shift depending on which dimensions of comparison (in the above cases, time versus money and positive versus negative features, respectively) are made more salient by the elicitation method. This calls into question whether subjects' preferences obey asymmetry, which requires that if *x* is strictly preferred to *y* then *y* cannot be strictly preferred to *x*. A more psychologically-derived view of people's preferences is that they are *constructed* when a preference question is asked, based on available cues and what comes to mind, rather than existing in a *stable* form prior to the inquiry, at least in a particular context *C*. The idea that preferences are reliable in *C* once an outcome has been specified is an important feature of the neoclassical view of human beings known as *Homo economicus* ("economic man"). In economic theories based on this view, preferences are often specified *exogenously* (existing outside the model). A view of preferences that says they are constructed supports descriptive models in which preferences are *endogenous* (derived within the model itself based on the context).

Much evidence from social psychology (e.g. the foot-in-the-door technique, the door-in-the-face technique, social proof, etc.) indicates that people construct their preferences rather than having a stable set.)

DEFINITION 3.1.8. If $P \subseteq \boldsymbol{X}\times\boldsymbol{X}$ is a strict preference relation, then the relation *I* (*indifference*) holds between *x* and *y* iff $x\neg Py$ and $y\neg Px$.

EXERCISE 3.1.9. Prove that if *P* is a preference relation that defines *I* as an indifference relation, then *I* is transitive.

DEFINITION 3.1.10. A strict preference relation P on an outcome set X is *representable in utility* iff there exists a $u: X \rightarrow \Re$ (function u that maps X into the real numbers) such that xPy iff $u(x) > u(y)$.

DEFINITION 3.1.11. A set $Y \subset X$ is *order-dense* under a preference relation P iff for all x and z in X, if xPz then there exists a y in Y such that xPy and yPz.

THEOREM 3.1.12. *Utility representation theorem for ordinal preferences*. A preference relation P on an outcome set X is representable in utility iff there exists a denumerable order-dense subset of X under P.
For the proof, see Blume (2006).

Theorem 3.1.12 is called a "representation theorem" because it specifies the conditions under which data (in this case preferences) can be represented by a function. The details of the theorem are beyond the scope of this course, since it refers to the distinction between denumerable (countable) and uncountable sets, a topic requiring deeper study of set theory than we will assume. But the defining condition for a preference relation to be representable in utility is an example of the *Archimedean condition*. Applied to preferences, an implication of the Archimedean condition is that it should never be possible for an arbitrarily small difference between outcomes on one dimension to outweigh an arbitrarily large difference on another dimension. Such a possibility can occur under what are called "lexicographic preferences".

DEFINITION 3.1.13. Suppose that two outcomes x and y may be compared along two dimensions d_1 and d_2. The agent *lexicographically prefers* x to y iff $d_1(x) > d_1(\mathrm{y})$ or $[d_1(x) = d_1(\mathrm{y})$ & $d_2(x) > d_2(\mathrm{y})]$. The definition can be extended for any higher number of dimensions.

Lexicographic preference relations are not representable in utility, and are often argued to be counternormative because they give infinite weight to infinitessimal differences.

EXERCISE 3.1.14. Prove that the lexicographic preference rule defined in 3.1.13 is a preference relation.

EXAMPLE 3.1.15. *Difference principle.* An example of a lexicographic preference rule that has been proposed seriously is the "difference principle" defined by John Rawls in *A Theory of Justice* (1971). The difference principle stipluates that any inequalities existing in a society "are to be of the greatest benefit to the least-advantaged members of society" (p. 303). This is also known as the *maximin* rule because it maximizes the minimum amount of resources that anyone in a society receives. The rule is lexicographic (and therefore violates the Archimedean condition) because any increase in resources for the least well-off person outweighs all changes in the resources received by others. A gain of one grain of rice by the least well-off can, under a strict interpretation of the difference principle, justify preferring that state of affairs to one in which everyone but the least well off would have vastly greater resources. This lexicographic character of the difference principle is less well known than the fact that Rawls's two main principles of justice are also lexicographically ordered (maximum equality of liberty takes lexicographic priority over economic well-being) and that the difference principle is lexicographically preferred to fair equality of opportunity.

Many studies of public opinion (e.g. Ferris, 1985) indicate that people hold lexicographic principles in ethical matters. For example, people often say that “human life comes before anything else” (Baron, 2000). It is difficult to determine experimentally how stringently people hold to these preferences when extreme costs on other dimensions are the result, but there is at least substantial evidence that professed policies adhere to lexicographic priority. Indeed, as shown above, philosophers have sometimes become famous for advocating lexicographic rules. Another example is Immanuel Kant, who took the view that dishonesty is always wrong, no matter what the consequences.

EXERCISE 3.1.16. Design an experiment to test whether people lexicographically prefer human life over some other factor, such as animal life.

3.2 Belief Neutrality

Consider two propositions, *p* and *q*. Each proposition can be treated as either an object of belief (e.g. *B(p,C)*, denoting that an agent believes *p* in context *C*) or as an object of preference (e.g. *pPq*). In the latter case, we will say that the agent prefers *p* being true over *q* being true in a context, where the truth of the proposition is itself treated as an outcome.

DEFINITION 3.2.1. *Polarity determination.* A relation *R determines* whether or not a relation *Q* holds across contexts *c* iff $[\forall c\ R(.;c) \Rightarrow Q(.;c)] \vee [\forall c\ R(.;c) \Rightarrow \neg Q(.;c)]$, where “.” refers to the set of predicated objects other than *c* for each relation. Quantification may also occur for other variables predicated upon by the relations *R* and *Q* (Davies, 1985).

DEFINITION 3.2.2. *Belief neutrality.* An agent's beliefs with respect to two propositions *p* and *q* in a context *C* are *neutral* with respect to the agent's preferences iff it is not the case that for the agent, *pPq* determines across contexts whether or not *Bp* or whether or not *Bq*.

Belief neutrality formalizes a principle that is often taken to be normative. If we assume that beliefs should be veridical (true), and that the truth is independent of what an agent prefers to be true, then preferences should not determine beliefs, i.e. belief formation should be neutral.

EXPERIMENT 3.2.3. *Self-deception*. Quattrone and Tversky (1984, cited in Baron, 2000) had subjects take a cold pressor test, in which they held their arm in cold water until they could no longer tolerate it. Subjects were then told that recent medical studies indicated a relationship between the test and two types of hearts: one associated with longer life and fewer heart attacks than the other type. Half the subjects were told that exercise (e.g. riding a stationary bicycle) increases tolerance for cold water for the good type of heart, while the other half were told that exercise decreases cold tolerance for the good type of heart. Subjects took the cold pressor test again after riding an exercycle for one minute. Subjects' tolerance for cold water tended to change in the direction consistent with what they had been told indicated a good heart, e.g. subjects told that the good heart increased tolerance after exercise held their arm in the cold water longer in the post-test. Only 9 out of 38 subjects stated anonymously afterward that they had purposely tried to change their tolerance in the direction associated with the good heart, although the remaining subjects showed just as large an effect. The 9 who admitted “cheating” also said they did not believe they actually had a good heart, but the 29 who said they had not cheated did believe it. This indicates that the 29 “non-cheaters” deceived themselves into believing what they preferred to be the case, namely that they had a good heart, thus violating belief neutrality.

3.3 Preferences and Contexts

DEFINITION 3.3.1. For *x,y* in an outcome set *X,* an agent's preference *xPy* is *context-independent* iff for any to subsets *Y* and Z of *X*, where *Y* and *Z* represent all of the *available* outcomes in different contexts, then if *x* and *y* are each elements of both *Y* and *Z*, then *xPy* when *Y* is the available set and *xPy* when *Z* is the available set. This principle has also been called *regularity* (Tversky and Simonson, 1993).

THEOREM 3.3.2. If *P* is a preference relation under an outcome set *X* that is representable in utility, then *P* is context independent for all *x* and *y* in *X*.
Proof. The utility function *u* that represents *P* does not depend on which set of outcomes is available within *X*. Thus, if *xPy* in a context in which *x* and *y* are in an available set $Y \subseteq X$, then $u(x) > u(y)$, which will be true if the available set changes to *Z*, so *xPy* when the available set is *Z*.

Context independence has been proposed as a normative principle on the grounds that the context of other available alternatives is irrelevant to a preference between two outcomes. Adding more outcomes to an available set might result in a change in the most preferred outcome if it had not previously been available, but it should not change how one ranks outcomes that were previously available. However, because people's preferences are often constructed in a context, and additional available options can affect perceptions of the relative merits of two options, people often violate context independence.

EXPERIMENT 3.3.3. *Context dependent preferences.* From Tversky and Simonson (1993): “In another study, subjects received descriptions and pictures of microwave ovens taken from the Best catalog. One group (n=60) was asked to choose between an Emerson priced at $110 and a Panasonic priced at $180. Both items were on sale, a third off the regular price. Here, 57% chose the Emerson and 43% chose the Panasonic. A second group (n=60) was presented with these options, along with a $200 Panasonic at a 10% discount. Because the two Panasonics were quite similar, the one with the lower discount appeared inferior to the other Panasonic, but it was not clearly inferior to the Emerson. Indeed, only 13% of the subjects chose the more expensive Panasonic, but its presence increased the percentage of subjects who chose the less expensive Panasonic from 43% to 60%, contrary to regularity.”