Choice among Similar Alternatives Revisited

Thomas J. Steenburgh

19 October 2007

What is IPS?

Tragically, it stands for the "Invariant Proportion of Substitution"

Newly discovered property that amounts to an implicit assumption about individual choice behavior

Like IIA:

- IPS leads to objectionable behavior in the context of choice among similar alternatives (as well as other contexts)
- Therefore, it is an undesirable property
- And, it might be useful to develop models that do not possess IPS

…But models that "fix" IIA do not necessarily fix IPS

Agenda

Recall Debreu's critique

Introduce IPS

- Explain why "fixing" IIA doesn't necessarily address Debreu's broader concerns
- Speculate why IPS has been overlooked (aggregation paradox)
- Show that allowing for differences in tastes doesn't necessarily address Debreu's critique either
- Show what IPS implies in a marketing mix context

Discuss Potential Fixes

Future Research

Debreu (1960)

Beethoven-Debussy (reprised in Tversky, 1972)

Suppose…

- An individual wants to buy a CD
- She is equally likely to choose a Beethoven or a Debussy recording

$$\Pr\{B_1 | B_1, D\} = \Pr\{D | B_1, D\} = \tfrac{1}{2}$$

Beethoven-Debussy

Now suppose she encounters a second Beethoven recording that she likes as well as the first:

$$Pr\{B_1 | B_1, B_2\} = Pr\{B_2 \; | B_1, B_2\} = \tfrac{1}{2}$$

Question

If the individual were rational, how would she choose among all three recordings $\{B_1, B_2, D\}$?

Following Debreu, we would expect...

$$\left.\begin{array}{l} Pr(B_1|B_1, B_2, D) = \frac{1}{4} \\ Pr(B_2|B_1, B_2, D) = \frac{1}{4} \end{array}\right\} \quad Pr(B_1 \text{ or } B_2|B_1, B_2, D) = \frac{1}{2}$$

$$Pr(D|B_1, B_2, D) = \frac{1}{2}$$

Principle: Similar alternatives should "suffer" more than dissimilar ones when a new alternative is added to the choice set

But, IIA implies that...

If a new alternative were added to the choice set, it <u>must</u> draw demand from the existing alternatives in proportion to the choice probabilities – 50% from B_1, 50% from D

Which means:

$Pr(B_1|B_1, B_2, D) = 1/3$

$Pr(B_2|B_1, B_2, D) = 1/3$

$Pr(D|B_1, B_2, D) = 1/3$

This is objectionable because…

But, IIA implies that...

If a new alternative were added to the choice set, it would <u>have</u> <u>to</u> draw demand from the existing alternatives in proportion to the choice probabilities.

Which means:

$$\left.\begin{array}{l} Pr(B_1|B_1, B_2, D) = 1/3 \\ Pr(B_2|B_1, B_2, D) = 1/3 \end{array}\right\} \quad Pr(B_1 \text{ or } B_2|B_1, B_2, D) = 2/3$$

$$Pr(D|B_1, B_2, D) = 1/3$$

While the individual's choice fundamentally stays the same, the odds of choosing Beethoven have undesirably increased.

Well-known examples include

•Beethoven / Debussy	Debreu, 1960; Tversky, 1972
•Pony / Bicycle	Savage (Luce and Suppes, 1965)
•Red bus / Blue bus	McFadden, 1974

None explicitly deal with the alternatives' attributes

McFadden (1974)

- Placed Luce's logit model in an economic framework
- Explicitly incorporated utility from <u>attributes</u>
- Used the model to estimate transportation demand

$$u_j = x_j\beta + \varepsilon_j \quad j = 1, \ldots, J$$

u_j – total utility for alternative j

$x_j\beta$ – utility from observed attributes
ε_j – utility from unobserved factors

McFadden (1974)

Key assumptions:

1. The unobserved utilities of any two alternatives are independent

   $$\varepsilon_j \perp \varepsilon_k \quad \forall j \neq k$$

2. The unobserved utilities are independent of the alternatives' attributes

   $$\varepsilon_j \perp x_k \quad \forall j,k$$

3. (weak complementarity) The observed utility of each alternative depends on only its own attributes

   $$v_j = f\left(x_j\right) \quad (\text{e.g. } v_j = x_j\beta)$$

McFadden (1974)

Assuming ε_j~ iid extreme value, the choice probabilities are:

$$\Pr\{j\} = \frac{e^{x_j\beta}}{\sum_{k=1}^{J} e^{x_k\beta}}$$

But this means the logit model possesses the IIA property

Models motivated by a desire to alleviate IIA

Generalized Extreme Value Models

- Nested logit (Williams, 1977; McFadden, 1978; Daly and Zachary, 1978)

Covariance Probit

Elimination-by-Aspects

- Tversky (1972)
- Batley and Daly (2006) shows these models are equivalent to GEV models

How GEV and covariance probit models work

Fix IIA by allowing the unobserved component of utility to be correlated across alternatives.

Model:

$$u_j = x_j\beta + \varepsilon_j \quad j = 1, \ldots, J$$

$$\mathrm{Corr}\left(\varepsilon_j, \varepsilon_k\right) \neq 0$$

Furthermore, the correlation between alternatives does not depend on their attributes.

Relax Assumption 1

Key assumptions:

1. The unobserved utilities of any two alternatives are independent $\quad \varepsilon_j \perp \varepsilon_k \quad \forall j \neq k$

2. The unobserved utilities are independent of the alternatives' attributes $\quad \varepsilon_j \perp x_k \quad \forall j,k$

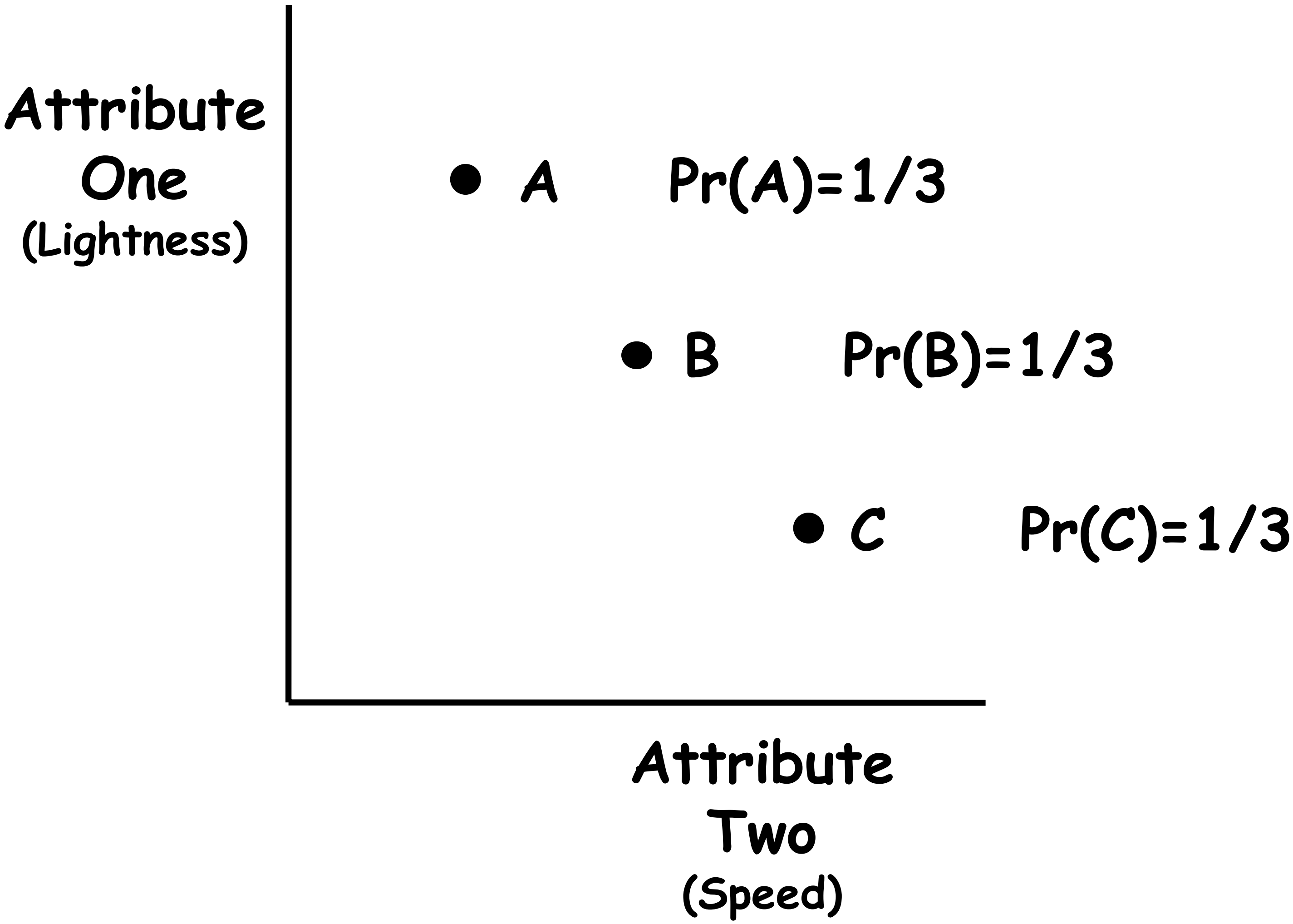
3. The observed utility of any alternative depends on only its own attributes $\quad v_j = f\left(x_j\right) \quad (\text{e.g. } v_j = x_j\beta)$

Do GEV and covariance probit models fully address Debreu's similarity critique?

Laptop Example

	Weight	Speed
Laptop A (Lightest)	3 lbs.	2.0 GHz
Laptop B	5 lbs.	2.7 GHz
Laptop C (Fastest)	7 lbs.	3.4 GHz

Consider an individual who is equally likely to choose any of the three laptops



In turn, let's now consider...

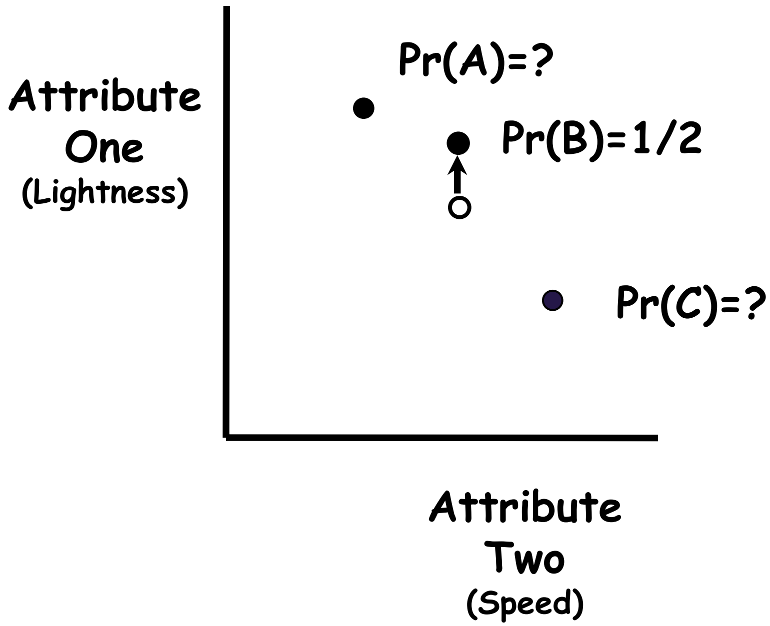
Two small improvements to Laptop B

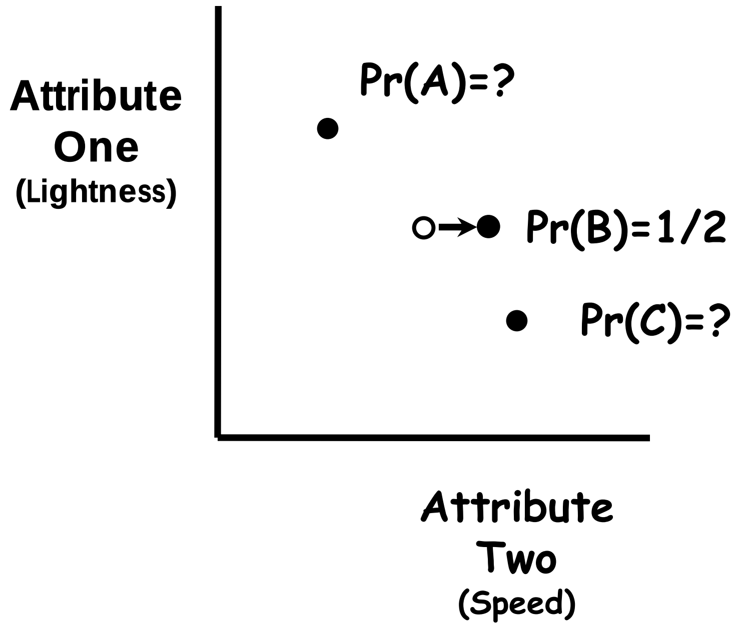
- One which makes it lighter
- Another which makes it faster

Suppose likelihood of its choice in both cases becomes $\frac{1}{2}$.

- which implies a growth in the likelihood of its choice of 1/6.

Question: How should the other alternatives suffer in either case?

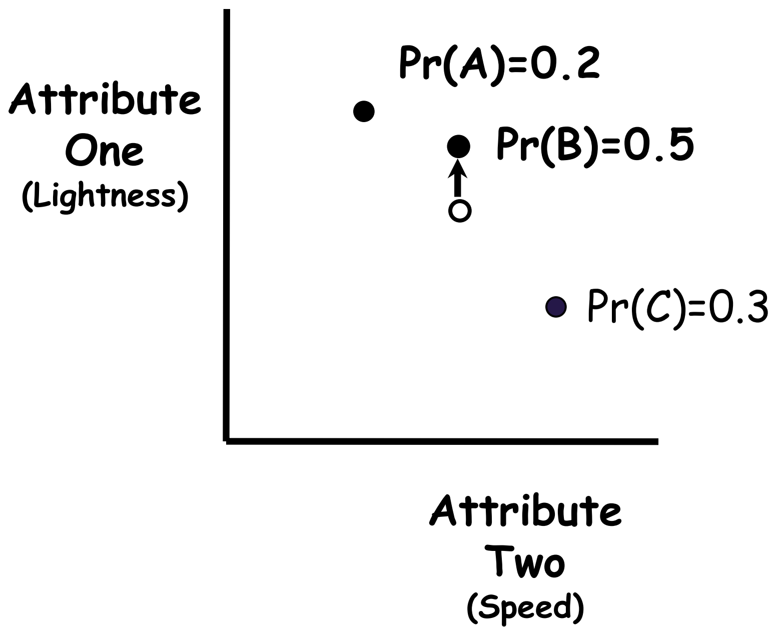


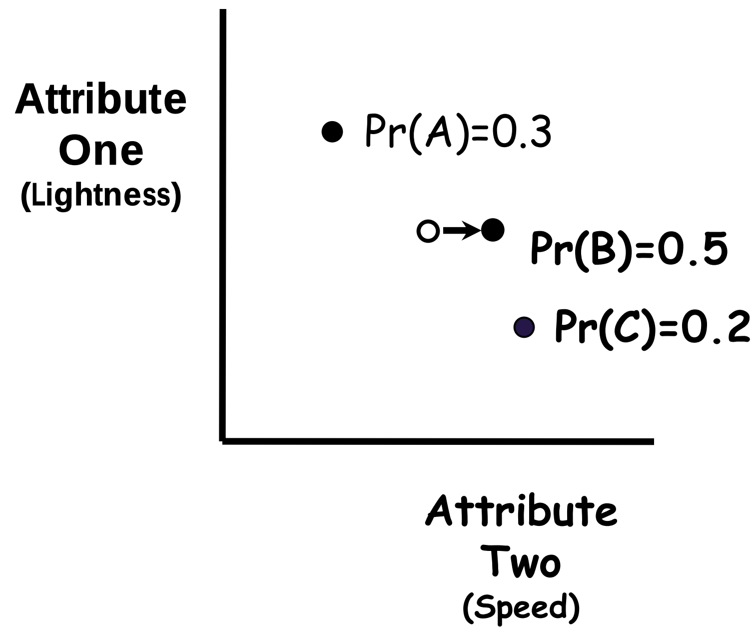


<u>Scenario one</u>
Improve B's weight

<u>Scenario two</u>
Improve B's speed

Given the similarity principle, we might expect...





Improving weight

a greater proportion drawn from A

Improving speed

a greater proportion drawn from C

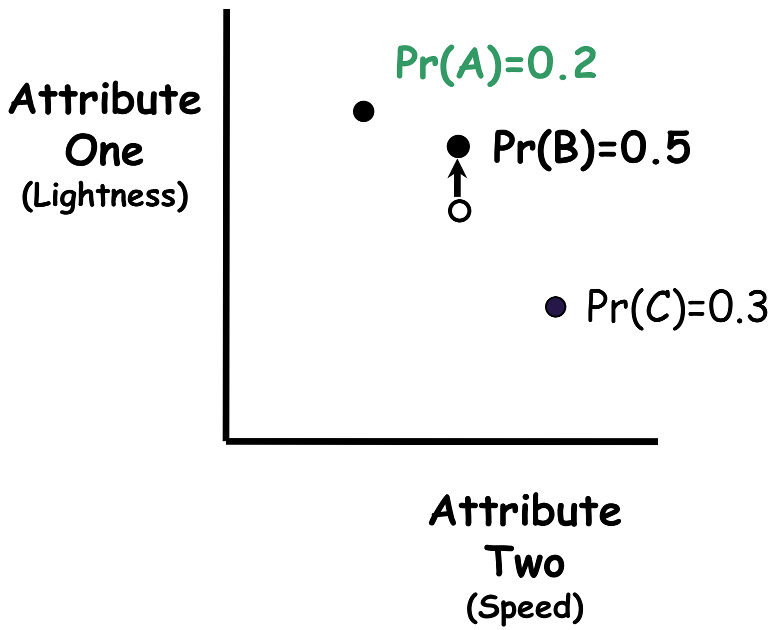
<u>But</u> IPS precludes this from happening

The proportion of growth in the likelihood of own-good choice drawn from a given competing alternative is the same no matter which attribute is improved.

$$\frac{-\partial P_k / \partial x_{ja}}{\partial P_j / \partial x_{ja}} = \Psi_{k/j} \qquad \forall a$$

IPS leads to a contradiction…

If the growth is mostly drawn from Laptop A...

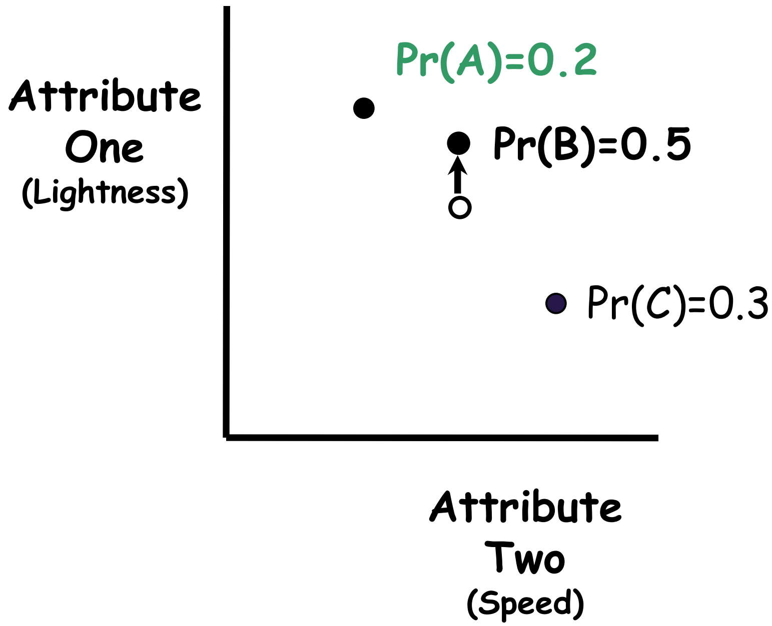


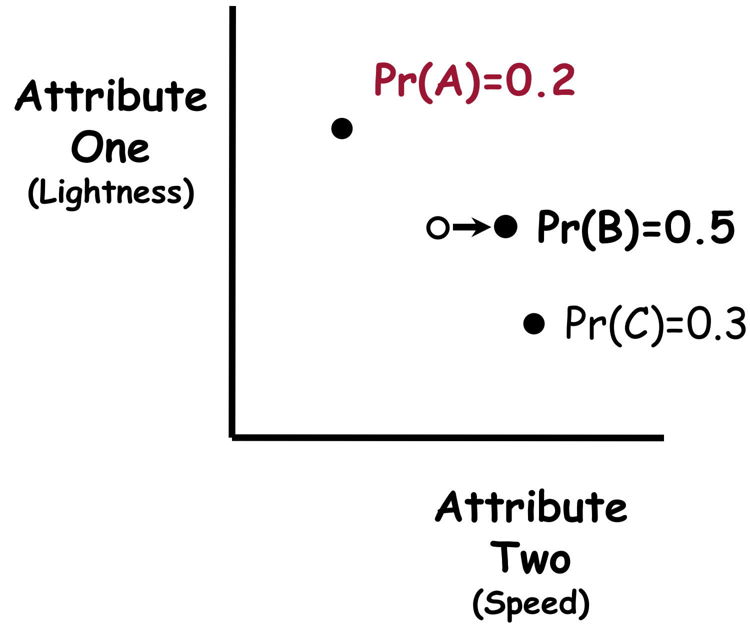
Say

$$\frac{-\partial P_A / \partial x_{Ba}}{\partial P_B / \partial x_{Ba}} = 0.8 \quad \forall a$$

<u>Improving weight</u>
reasonable to draw more from A

If the growth is mostly drawn from Laptop A…

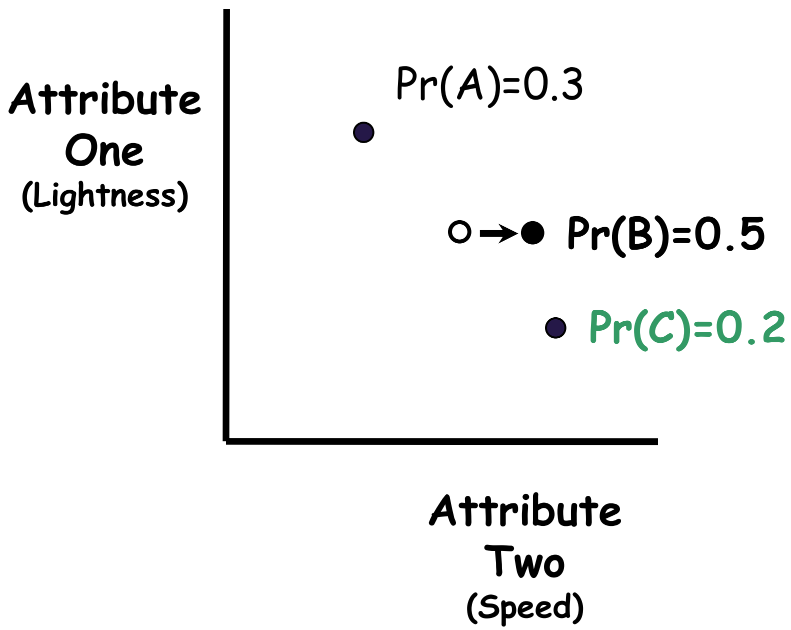




Improving weight
reasonable to draw more from A

Improving speed
counterintuitive to draw more from C

If the growth is mostly drawn from Laptop C...

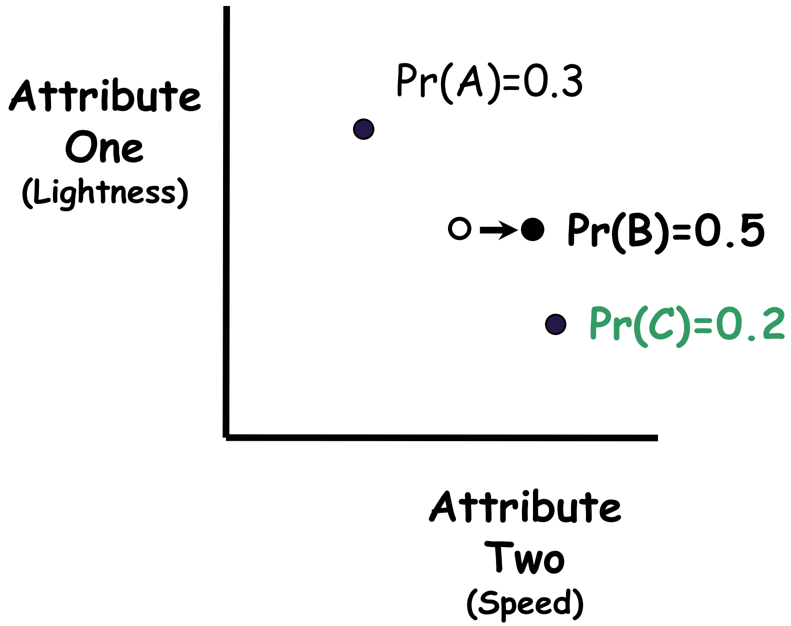


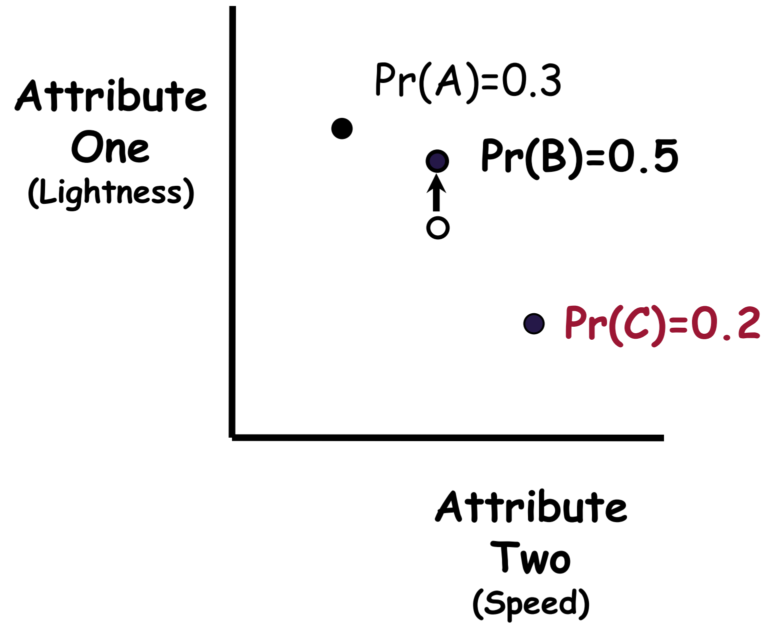
Say

$$\frac{-\partial P_C / \partial x_{Ba}}{\partial P_B / \partial x_{Ba}} = 0.8 \quad \forall a$$

Improving speed
reasonable to draw more from C

If the growth is mostly drawn from Laptop C...





Improving speed
reasonable to draw more from C

Improving weight
counterintuitive to draw more from C

The cool thing is...

- While GEV and covariance probit models may technically “fix” IIA, they still possess IPS
- **More importantly, these models do not completely address the similarity critique because they ignore the direction of change**
- **Why?** The correlation in unobserved utility between any two alternatives is fixed, so their “similarity” is fixed too

Why might have the IPS property been overlooked?

A possible reason...

IPS does not necessarily hold in aggregate even if it hold in every subpopulation

Consider two homogeneous customer segments, both of which substitute according to IPS

- Salespeople who value lighter weights more than faster speeds, say $Pr\{A|A,C\}=0.7$
- Scientists who value the opposite, say $Pr\{C|A,C\}=0.7$

Salespeople's substitution under IPS

	Improve B's Weight	Improve B's Speed
Laptop A (Lightest)	-14 units (70%)	-7 units (70%)
Laptop B	+20 units	+10 units
Laptop C (Fastest)	-6 units (30%)	-3 units (30%)

Scientists' substitution under IPS

	Improve B's Weight	Improve B's Speed
Laptop A (Lightest)	-3 units (30%)	-6 units (30%)
Laptop B	+10 units	+20 units
Laptop C (Fastest)	-7 units (70%)	-14 units (70%)

What happens in aggregate?

It's like Simpson's Paradox in reverse

- We object to the substitution behavior of both salespeople and scientists
- But, in aggregate, it looks like they're behaving reasonably

Aggregate substitution

	Improve B's Weight	Improve B's Speed
Laptop A (Lightest)	-17 units **(57%)**	-13 units (43%)
Laptop B	+30 units	+30 units
Laptop C (Fastest)	-13 units (43%)	-17 units **(57%)**

Aggregate substitution

Not only does IPS not technically hold in aggregate, but both…

- The proportion of demand drawn from Laptop A (the lightest alternative) is greater when Laptop B is made lighter as opposed to faster (57% vs. 43%)
- The proportion of demand drawn from Laptop C (the fastest alternative) is greater when Laptop B is made faster as opposed to lighter (57% vs. 43%)

which seems reasonable.

Still we might not be completely satisfied

Other, perhaps more reasonable, individual substitution behavior might better account for the aggregate substitution results.

For instance, the following individual substitution behavior…

Individual Substitution Precluded by IPS

	Salespeople		Scientists	
	Improve B's Weight	Improve B's Speed	Improve B's Weight	Improve B's Speed
Laptop A (Lightest)	-11.3 units **(57%)**	-4.3 units (43%)	-5.7 units **(57%)**	-8.7 units (43%)
Laptop B	+20 units	+10 units	+10 units	+20 units
Laptop C (Fastest)	-8.7 units (43%)	-5.7 units (57%)	-4.3 units (43%)	-11.3 units **(57%)**

Individual Substitution Precluded by IPS

	Salespeople		Scientists	
	Improve B's Weight	Improve B's Speed	Improve B's Weight	Improve B's Speed
Laptop A (Lightest)	-11.3 units (57%)	-4.3 units (43%)	-5.7 units (57%)	-8.7 units (43%)
Laptop B	+20 units	+10 units	+10 units	+20 units
Laptop C (Fastest)	-8.7 units (43%)	-5.7 units (57%)	-4.3 units (43%)	-11.3 units (57%)

Produces the same aggregate results

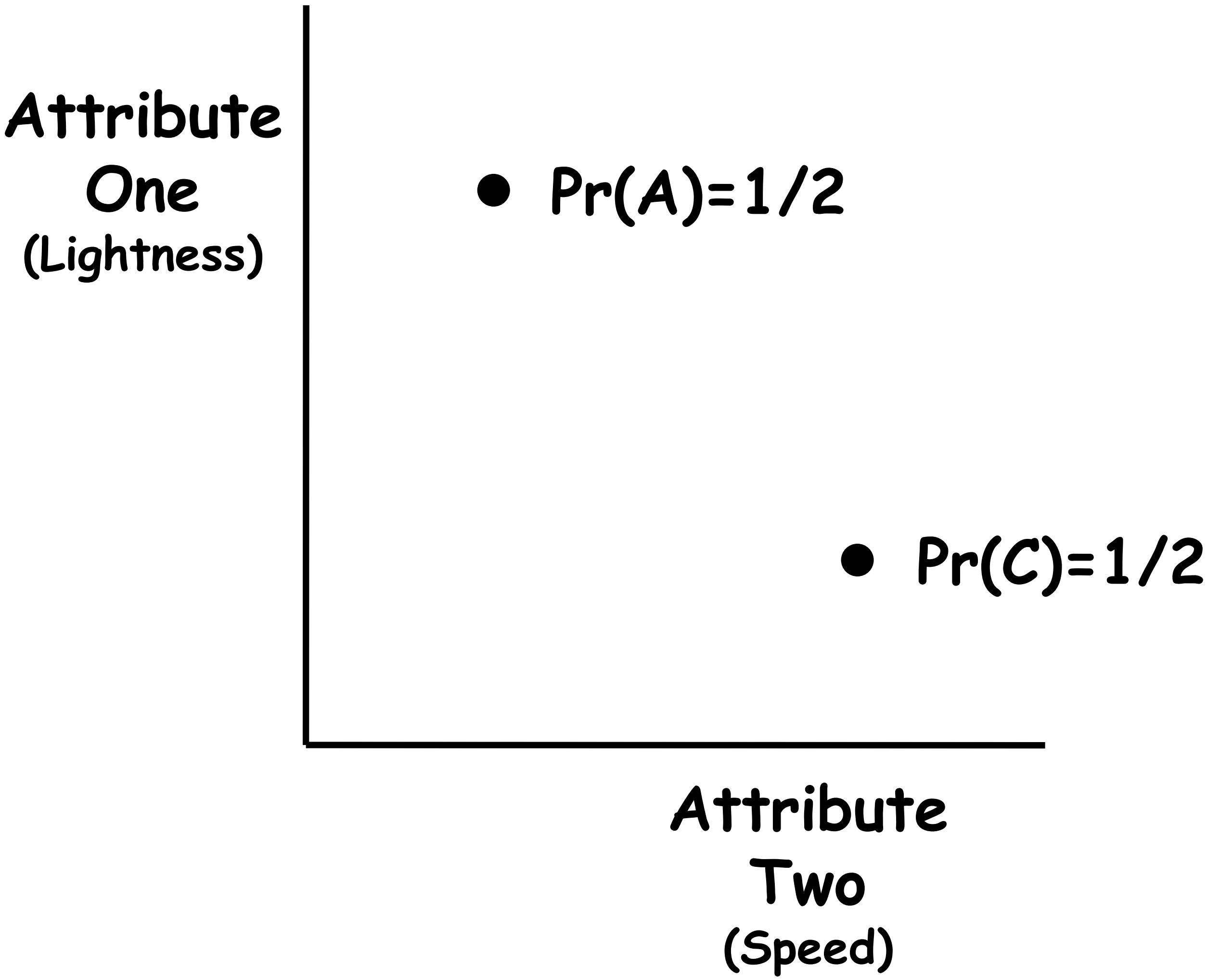
	Improve B's Weight	Improve B's Speed
Laptop A (Lightest)	-17 units **(57%)**	-13 units (43%)
Laptop B	+30 units	+30 units
Laptop C (Fastest)	-13 units (43%)	-17 units **(57%)**

What do we learn from this example?

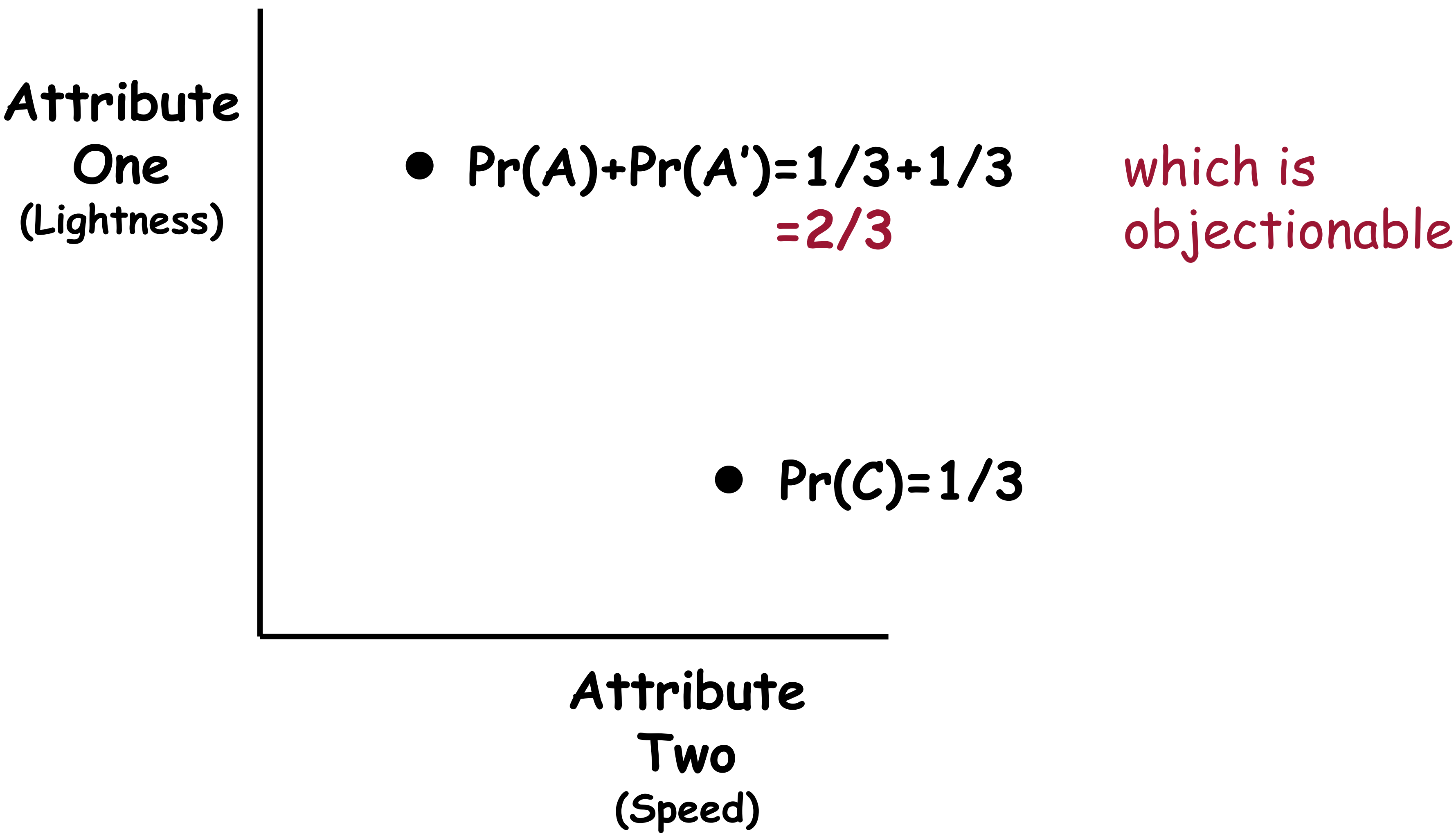
- Cannot draw conclusions about individual substitution behavior from the aggregate patterns predicted by the model
- There's an opportunity to construct better models of individual behavior to let us do that

Does allowing for differences in tastes (heterogeneity) fully address Debreu's similarity critique?

Consider an individual who is equally likely to choose either of two laptops



Recall under IIA, if we add an alternative identical to Laptop A...

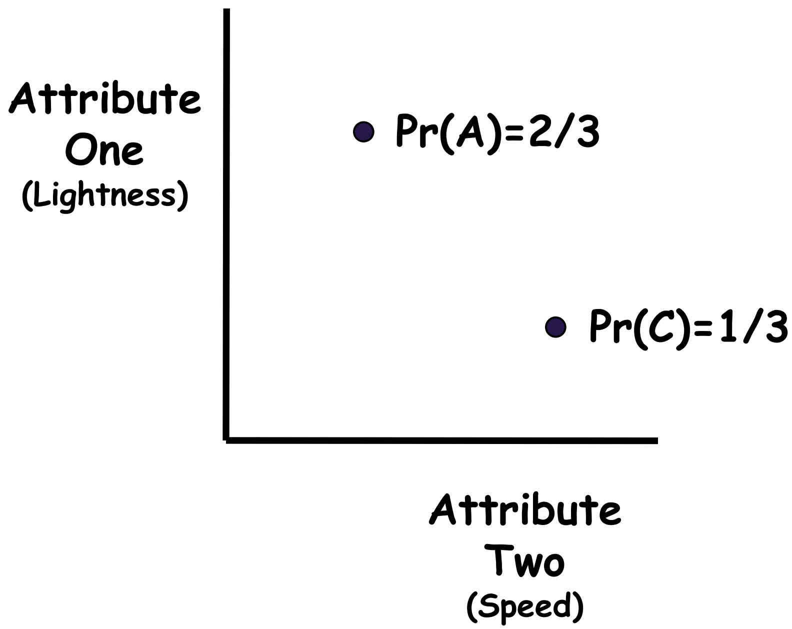


Counterexample

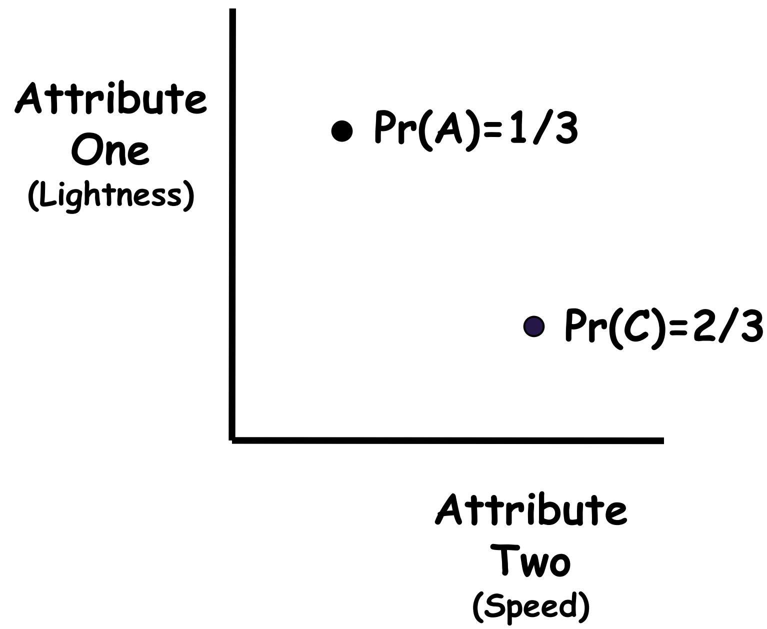
- Suppose we have two people with opposite preferences
- But an individual chosen at random from the population would be equally likely to choose either alternative

For instance, suppose...

Salespeople

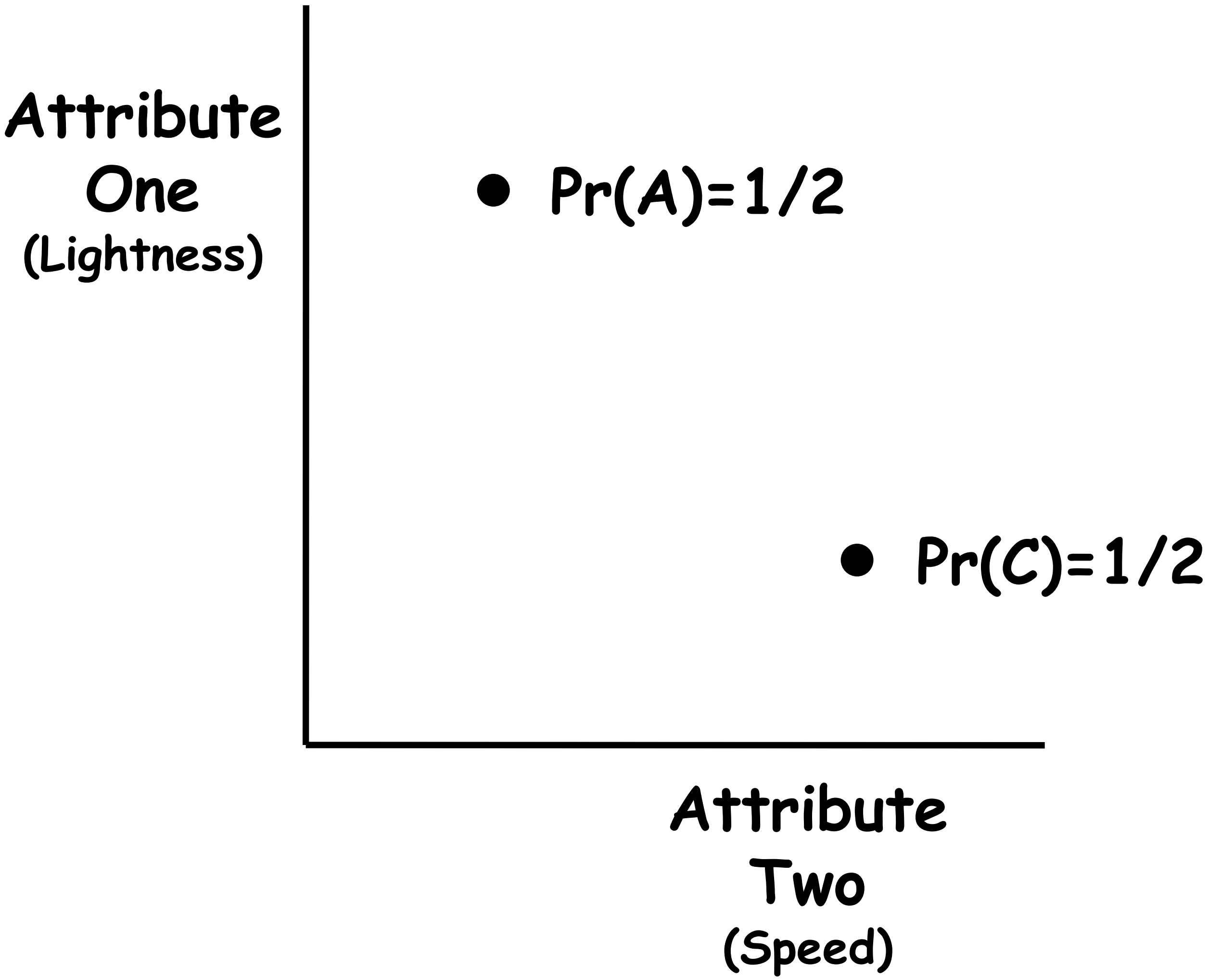


Scientists



Pr(Salesperson) = Pr(Scientist) = 1/2

Thus, a randomly chosen individual's preferences are...



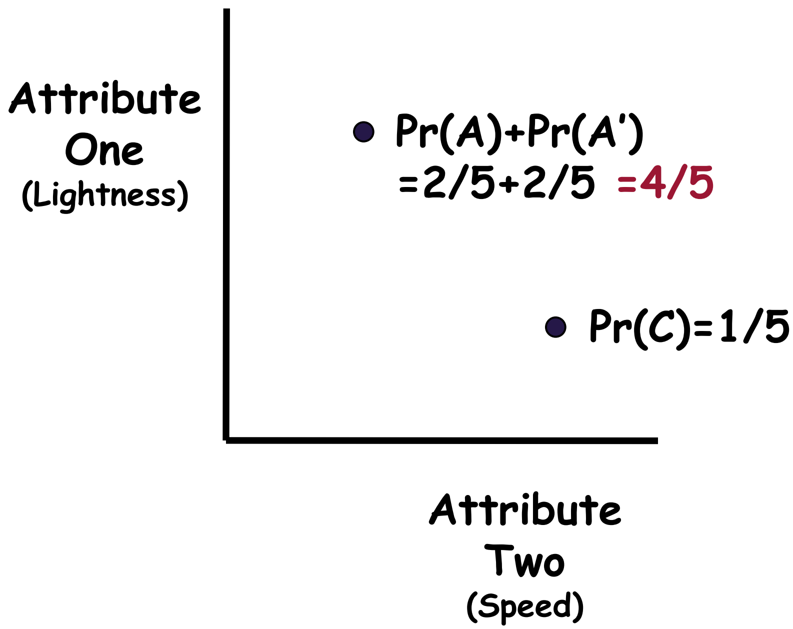
How would the representative individual substitute among laptops?

Suppose:

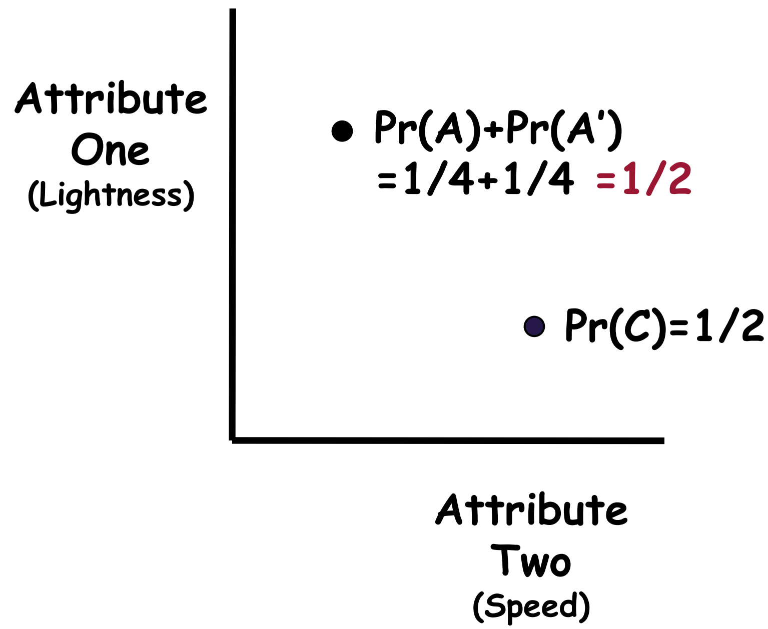
- *A* new alternative identical to Laptop *A*, call it *A′*, is introduced into the choice set, so {*A*,*A′*,*C*}
- Both salespeople and scientists substitute according to IIA

This results in

Salespeople



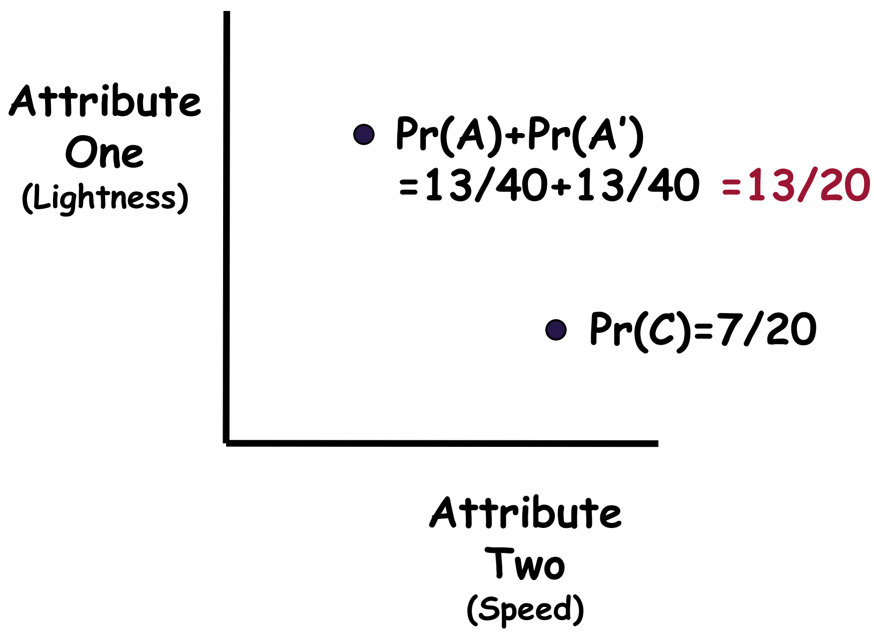
Scientists



which is objectionable for each individual.

Does the similarity critique still apply? Yes!

Representative Individual



The objection isn't quite as great:

- In this case, Pr(A or A') is 13/20 ≈ 65%
- Whereas with homogeneous tastes, Pr(A or A') = 2/3

But the representative individual chooses the lighter laptop with greater probability

Where might IPS matter?

Where might IPS Matter?

Contexts where the similarity changes:

- Conjoint experiments
- Markets in which goods compete by changing attributes
- Behavioral studies in which a "rational" model is treated as a baseline

Could also matter in marketing mix studies.

Consider a marketing mix problem

Two goods are competing in a market

- Cabot Farms milk
- Land o' Lakes milk

Cabot can either drop price or invest in advertising

Marketing mix model

$$u_j = price_j \beta_1 + ad_j \beta_2 + \varepsilon_j \quad j = 1, \ldots, J$$

The nested logit, a GEV model, assumes fixed correlation of ε among alternatives

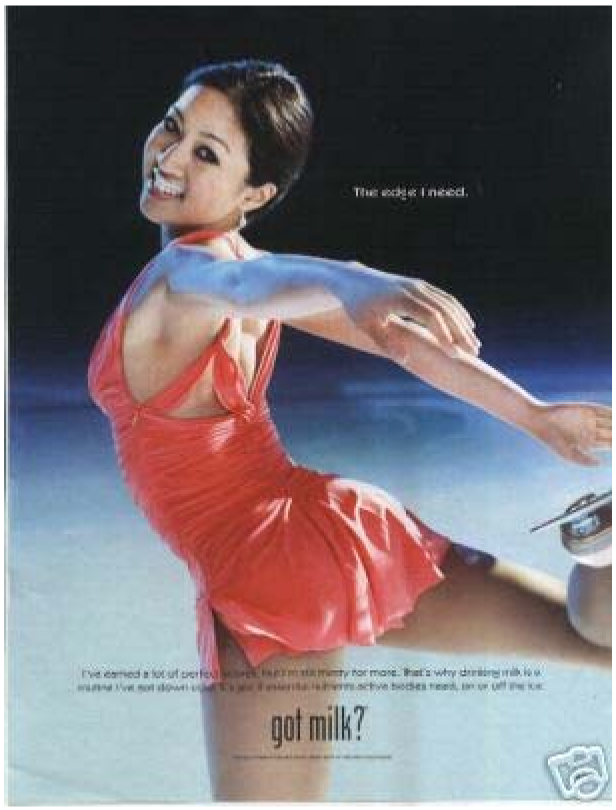
Where should demand be drawn if Cabot

- Drops price?
- Invests in awareness building advertising?

If Cabot drops price, we might expect...

- Some demand to come from consumers entering the market (say 75%)
- The rest comes from brand switching (say 25%)

But if Cabot invests in “Got Milk?” ads?



- All of Cabot’s demand would come from consumers entering the market
- In fact, we might expect this ad to create demand for Land O’Lakes too

The following unit-based decomposition might seem reasonable...

	Drop price	Increase ads
Cabots	+100 units	+80 units
Land O'Lakes	-25 units (25%)	+80 units (-100%)
Outside good	-75 units (75%)	-160 units (200%)

...but these substitution patters would be precluded by IPS.

Because under IPS, if...

	Drop price	Increase ads
Cabots	+100 units	+80 units
Land O'Lakes	-25 units (25%)	
Outside good	-75 units (75%)	

Then…

	Drop price	Increase ads
Cabots	+100 units	+80 units
Land O'Lakes	-25 units (25%)	-20 units (25%)
Outside good	-75 units (75%)	

…the proportion of demand drawn from Land O'Lakes has to be the same across all marketing instruments.

Then…

	Drop price	Increase ads
Cabots	+100 units	+80 units
Land O'Lakes	-25 units (25%)	-20 units (25%)
Outside good	-75 units (75%)	-60 units (75%)

…the proportion of demand drawn from the outside good has to be the same across all marketing instruments.

How might the discrete-choice models be generalized if we want to solve these problems?

Context of choice among similar alternatives

Error-components model:

$$u_j = v\left(x_j\right) + \varepsilon_j\left(x_j\right)$$

- The unobserved utilities, $\varepsilon_j(x_j)$, are functions of the alternatives' attributes
- The unobserved utilities of any two alternatives become more correlated as the alternatives become more similar
- This challenges the idea of unobserved utility though

Potential Fixes

Key assumptions:

1. The unobserved utilities of any two alternatives are independent $\qquad \varepsilon_j \perp \varepsilon_k \quad \forall j \neq k$
2. The unobserved utilities are independent of the alternatives' attributes $\qquad \varepsilon_j \perp x_k \quad \forall j, k$
3. The observed utility of any alternative depends on only its own attributes $\qquad v_j = f(x_j) \quad (\text{e.g. } v_j = x_j\beta)$

Marketing mix context

Universal logit model:

$$u_j = v_j\left(x_1,\ldots,x_J\right) + \varepsilon_j$$

- The indirect utilities, v_j, depend on the attributes of competing alternatives
- This challenges the notion of utility
- Economists have questioned whether this is a RUM

Potential Fixes

Key assumptions:

1. The unobserved utilities of any two alternatives are independent $\quad \varepsilon_j \perp \varepsilon_k \quad \forall j \neq k$
2. The unobserved utilities are independent of the alternatives' attributes $\quad \varepsilon_j \perp x_k \quad \forall j,k$
3. The observed utility of any alternative depends on only its own attributes $\quad v_j = f(x_j) \quad (e.g.\ v_j = x_j\beta)$

What I would like you to remember

- Debreu's critique is not overcome simply by fixing IIA
- IPS is a property that *gives* us insight about how the discrete-choice models work and may lead to more useful models
- These advances may be found with clearer depictions of individual choice behavior
  (as opposed to aggregate models that simply fit the data well)