Sequential Approval

Paola Manzini    Marco Mariotti    Levent Ülkü
Motivation

Consider an individual who:

Scans headlines and reads some articles (or stores away for later reading)
Consider an individual who:

Scans headlines and reads some articles (or stores away for later reading)

‘Likes’ or shares various posts when going down his social media feed
Consider an individual who:

Scans headlines and reads some articles (or stores away for later reading)

‘Likes’ or shares various posts when going down his social media feed

Progressively fills her online shopping cart
Consider an individual who:

Scans headlines and reads some articles (or stores away for later reading)

‘Likes’ or shares various posts when going down his social media feed

Progressively fills her online shopping cart

Progressively ‘matches’ with potential partners on an online dating site
These are diverse examples of behaviour. What do they have in common?
These are diverse examples of behaviour. What do they have in common?

i) Objects come \textit{sequentially} to the attention of the agent: they form a \textit{list}.

- There’s not going to be a final choice between articles read or posts Liked or shared.
- An item/partner may or may not be finally selected from the cart/match set.
- The whole cart/match set may even be abandoned (they act as ‘consideration sets’).

- The order aspect overrides the menu aspect.
- Some capacity constraint is likely to apply (cannot go on forever).
These are diverse examples of behaviour. What do they have in common?

i) Objects come *sequentially* to the attention of the agent: they form a *list*

ii) Objects are not quite ‘chosen’, but merely ‘approved’:
- there’s not going to be a final choice between articles read or posts Liked or shared.
- an item/partner may or may not be finally selected from the cart/match set
- the whole cart/match set may even be abandoned (they act as ‘consideration sets’)

Manzini Mariotti Ülkü
Sequential Approval
BRIColumbia

These are diverse examples of behaviour. What do they have in common?

i) Objects come *sequentially* to the attention of the agent: they form a *list*

ii) Objects are not quite ‘chosen’, but merely ‘approved’:
- there’s not going to be a final choice between articles read or posts Liked or shared.
- an item/partner may or may not be finally selected from the cart/match set
- the whole cart/match set may even be abandoned (they act as ‘consideration sets’)

iii) The set of potentially approvable objects is very large, even ‘endless’, so:
- the *order* aspect overrides the *menu* aspect
- some *capacity constraint* is likely to apply (cannot go on forever)
Build a model of sequential approval with features i-iii:

$X$ is a finite set of alternatives, $n$ its cardinality (thought of as being very large) $X$, and $N = \{1, \ldots, n\}$.

A list is any linear order $\lambda$ on $X$, sometimes denoted $xyz$...

$\Lambda$ is the set of all lists.

A stochastic approval function is a map $p : X \times \Lambda \to [0, 1]$.

The number $p(x, \lambda)$ is the probability that $x$ is approved when the decision maker is facing list $\lambda$. 
Have we forgotten the adding up constraint on probabilities?
Sequential approval - ctnd.

Have we forgotten the adding up constraint on probabilities?
No: the sum of the $p(x, \lambda)$ over alternatives is typically not one. This is approval, not choice...
Have we forgotten the adding up constraint on probabilities?

No: the sum of the $p(x, \lambda)$ over alternatives is typically not one. This is approval, not choice...

Formally, a stochastic approval function could be equivalently defined as a stochastic correspondence $C : 2^X \times \Lambda \rightarrow [0, 1]$ associating with each list the probability of the possible approval sets.

The adding up constraint applies over these objects, provided that $C(\emptyset, \lambda)$ is allowed to absorb residual probability.
Have we forgotten the adding up constraint on probabilities?

No: the sum of the $p(x, \lambda)$ over alternatives is typically not one. This is approval, not choice...

Formally, a stochastic approval function could be equivalently defined as a stochastic correspondence $C : 2^X \times \Lambda \rightarrow [0, 1]$ associating with each list the probability of the possible approval sets.

The adding up constraint applies over these objects, provided that $C (\emptyset, \lambda)$ is allowed to absorb residual probability.

Unlike a standard stochastic choice correspondence, our domain comprises lists, not menus. The menu $X$ is held fixed in the analysis. The variation comes only from lists.
At the moment of approving the agent is defined by:

1) A *preference* \( \succeq \) (a linear order over \( X \)).

2) An *approval threshold* (an element of \( X \)).

3) A *stopping rule* (a number expressing a capacity constraint - more on this later).
We take preference to be the stable element of the agent’s psychology. Approval thresholds and capacity constraints are subject to random shocks.

E.g:
- each morning you may be more or less strict with your FB Likes
- each morning you may have more or less time/patience to go down the list.
Let $\pi$ be a (strictly positive) joint probability distribution over $N \times X$ that describes this randomness.

$\pi(i, t)$ is the joint probability that the approval threshold is $t$ and the capacity constraint is $i$. An agent is a pair $(\succeq, \pi)$. 

Let $\pi$ be a (strictly positive) joint probability distribution over $N \times X$ that describes this randomness.

$\pi(i, t)$ is the joint probability that the approval threshold is $t$ and the capacity constraint is $i$. An agent is a pair $(\succeq, \pi)$.

Consider two possibilities regarding the capacity constraint:

i) *Depth constraint*: the constraint acts on the number of alternatives you examine.

ii) *Approval constraint*: the constraint acts on the number of alternatives you approve.
Let $\lambda(x)$ be the position of $x$ in list $\lambda$.

The DCM is represented as

$$ p^D(x, \lambda) = \sum_{x \geq t} \sum_{i \geq \lambda(x)} \pi(i, t) $$
Let $\lambda(x)$ be the position of $x$ in list $\lambda$.

The DCM is represented as

$$p^D(x, \lambda) = \sum_{x \geq t} \sum_{i \geq \lambda(x)} \pi(i, t)$$

Let $b(\lambda, j, t)$ be the number of alternatives that are $\geq t$ and that in list $\lambda$ are in a position $\leq j$.

The ACM is represented as

$$p^A(x, \lambda) = \sum_{x \geq t} \sum_{i \geq b(\lambda, \lambda(x), t)} \pi(i, t)$$
- Rubinstein & Salant TE06 (mostly observable lists, menu variation, choice functions - or correspondences by taking unions of lists)
- Yildiz TE16 (menu variation, rationalisation by lists)
- Aguiar, Boccardi & Dean JET16 (menu variation, rationalisation by random lists)
- Kovach & Ülkü 2017 (menu variation, rationalisation by lists, random threshold).
- Caplin, Dean & Martin AER11 (experimental choice process data, infer search order).
Our Questions

1) Identification:

2) List design:

3) Characterisation:

4) Comparative statics:
Our Questions

1) **Identification**: Assume that approvals are generated by the model(s). Can an observer of approvals and lists identify the parameters, i.e. preferences $\succ$ and the joint probabilities $\pi(i, t)$?

2) **List design**:

3) **Characterisation**:

4) **Comparative statics**:
1) Identification: Assume that approvals are generated by the model(s). Can an observer of approvals and lists identify the parameters, i.e. preferences $\succeq$ and the joint probabilities $\pi(i, t)$?

2) List design: Given an objective (e.g. total number of approvals), which list maximises the objective?

3) Characterisation:

4) Comparative statics:
Our Questions

1) *Identification*: Assume that approvals are generated by the model(s). Can an observer of approvals and lists identify the parameters, i.e. preferences \( \succeq \) and the joint probabilities \( \pi(i, t) \)?

2) *List design*: Given an objective (e.g. total number of approvals), which list maximises the objective?

3) *Characterisation*: Which exact constraints on observed approval behaviour do the models impose?

4) *Comparative statics*:
1) *Identification*: Assume that approvals are generated by the model(s). Can an observer of approvals and lists identify the parameters, i.e. preferences $\succeq$ and the joint probabilities $\pi(i, t)$?

2) *List design*: Given an objective (e.g. total number of approvals), which list maximises the objective?

3) *Characterisation*: Which exact constraints on observed approval behaviour do the models impose?

4) *Comparative statics*: How are changes in the primitives manifested in behaviour?
Let $x \succ y \succ z$. Denote the marginals of $\pi$ by $\pi_i$ and $\pi_t$.

<table>
<thead>
<tr>
<th>DCM</th>
<th>$p(x, \lambda)$</th>
<th>$p(y, \lambda)$</th>
<th>$p(z, \lambda)$</th>
</tr>
</thead>
</table>

Only the position in the list matters.
3-Alternative example: DCM

Let $x \succ y \succ z$. Denote the marginals of $\pi$ by $\pi_i$ and $\pi_t$.

<table>
<thead>
<tr>
<th>DCM</th>
<th>$p(x, \lambda)$</th>
<th>$p(y, \lambda)$</th>
<th>$p(z, \lambda)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = xyz$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Only the position in the list matters.
Let $x \succ y \succ z$. Denote the marginals of $\pi$ by $\pi_i$ and $\pi_t$.

<table>
<thead>
<tr>
<th>DCM</th>
<th>( p(x, \lambda) )</th>
<th>( p(y, \lambda) )</th>
<th>( p(z, \lambda) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda = xyz )</td>
<td>( \sum_{i=1}^{3} \sum_{t \in {x, y, z}} \pi_{i,t} = 1 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Let $x \succ y \succ z$. Denote the marginals of $\pi$ by $\pi_i$ and $\pi_t$.

<table>
<thead>
<tr>
<th>DCM</th>
<th>$p(x, \lambda)$</th>
<th>$p(y, \lambda)$</th>
<th>$p(z, \lambda)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = xyz$</td>
<td>$\frac{3}{3} \sum_{i=1}^{3} \sum_{t \in {x, y, z}} \pi_{i,t} = 1$</td>
<td>$\frac{3}{3} \sum_{i=2}^{3} \sum_{t \in {y, z}} \pi_{i,t}$</td>
<td>$\frac{3}{3} \sum_{i=2}^{3} \sum_{t \in {y, z}} \pi_{i,t}$</td>
</tr>
</tbody>
</table>

Only the position in the list matters.
3-Alternative example: DCM

Let $x \succ y \succ z$. Denote the marginals of $\pi$ by $\pi_i$ and $\pi_t$.

<table>
<thead>
<tr>
<th>DCM</th>
<th>$p(x, \lambda)$</th>
<th>$p(y, \lambda)$</th>
<th>$p(z, \lambda)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = xyz$</td>
<td>$\sum_{i=1}^{3} \sum_{t \in {x,y,z}} \pi_{i,t} = 1$</td>
<td>$\sum_{i=2}^{3} \sum_{t \in {y,z}} \pi_{i,t}$</td>
<td>$\pi_{3,z}$</td>
</tr>
</tbody>
</table>
3-Alternative example: DCM

Let $x \succ y \succ z$. Denote the marginals of $\pi$ by $\pi_i$ and $\pi_t$.

<table>
<thead>
<tr>
<th>DCM</th>
<th>$p(x, \lambda)$</th>
<th>$p(y, \lambda)$</th>
<th>$p(z, \lambda)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = xyz$</td>
<td>$\sum_{i=1 \atop t \in {x, y, z}}^3 \pi_{i,t} = 1$</td>
<td>$\sum_{i=2 \atop t \in {y, z}}^3 \pi_{i,t}$</td>
<td>$\pi_{3,z}$</td>
</tr>
<tr>
<td>$\lambda = xzy$</td>
<td>$\sum_{i=1 \atop t \in {x, y, z}}^3 \pi_{i,t} = 1$</td>
<td>$\sum_{t \in {y, z}} \pi_{3,t}$</td>
<td>$\sum_{i=2}^3 \pi_{i,z}$</td>
</tr>
<tr>
<td>$\lambda = yxz$</td>
<td>$\sum_{i=2}^3 \pi_i$</td>
<td>$\pi_y + \pi_z$</td>
<td>$\pi_{3,z}$</td>
</tr>
<tr>
<td>$\lambda = yzx$</td>
<td>$\pi_3$</td>
<td>$\pi_y + \pi_z$</td>
<td>$\sum_{i=2}^3 \pi_{i,z}$</td>
</tr>
<tr>
<td>$\lambda = zxy$</td>
<td>$\sum_{i=2}^3 \pi_i$</td>
<td>$\sum_{t \in {y, z}} \pi_{3,t}$</td>
<td>$\pi_z$</td>
</tr>
<tr>
<td>$\lambda = zyx$</td>
<td>$\pi_3$</td>
<td>$\sum_{i=2 \atop t \in {y, z}}^3 \pi_{i,t}$</td>
<td>$\pi_z$</td>
</tr>
</tbody>
</table>

*Only the position in the list matters.*
### 3-Alternative example: ACM

<table>
<thead>
<tr>
<th>ACM</th>
<th>$p(x, \lambda)$</th>
<th>$p(y, \lambda)$</th>
<th>$p(z, \lambda)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = xyz$</td>
<td>(\sum_{i=1}^{3} \sum_{t \in {x,y,z}} \pi_{i,t} = 1)</td>
<td>(\sum_{i=2}^{3} \sum_{t \in {y,z}} \pi_{i,t})</td>
<td>$\pi_{3,z}$</td>
</tr>
<tr>
<td>$\lambda = xzy$</td>
<td>(\sum_{i=1}^{3} \sum_{t \in {x,y,z}} \pi_{i,t} = 1)</td>
<td>(\sum_{i=2}^{3} \pi_{i,y} + \pi_{3,z})</td>
<td>(\sum_{i=2}^{3} \pi_{i,z})</td>
</tr>
<tr>
<td>$\lambda = yxz$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Here predecessor set matters for approval probs, not just position.
### 3-Alternative example: ACM

<table>
<thead>
<tr>
<th>ACM</th>
<th>( p(x, \lambda) )</th>
<th>( p(y, \lambda) )</th>
<th>( p(z, \lambda) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda = xyz )</td>
<td>( \sum_{i=1}^{3} \sum_{t \in {x,y,z}} \pi_{i,t} = 1 )</td>
<td>( \sum_{i=2}^{3} \sum_{t \in {y,z}} \pi_{i,t} )</td>
<td>( \pi_{3,z} )</td>
</tr>
<tr>
<td>( \lambda = xzy )</td>
<td>( \sum_{i=1}^{3} \sum_{t \in {x,y,z}} \pi_{i,t} = 1 )</td>
<td>( \sum_{i=2}^{3} \pi_{i,y} + \pi_{3,z} )</td>
<td>( \sum_{i=2}^{3} \pi_{i,z} )</td>
</tr>
<tr>
<td>( \lambda = yxz )</td>
<td>( \pi_{1,x} + \sum_{i=2}^{3} \pi_{i} )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Here *predecessor set matters* for approval probs, not just position.
The identification news regarding DCM is good:
The identification news regarding DCM is good:

**Theorem**

*In the DCM, preferences and joint probabilities $\pi(i, t)$ are uniquely identified by approval probabilities.*
The identification news regarding DCM is good:

**Theorem**

*In the DCM, preferences and joint probabilities \( \pi(i, t) \) are uniquely identified by approval probabilities.*

The result hinges strongly on the fact that here the position of an alternative in a list determines the approval probability.
Proof (sketch)

Preferences between any two alternatives \( x \) and \( y \) are identified by the ranking of approval probabilities in any lists in which \( x \) and \( y \) are in the same position (see example).
Preferences between any two alternatives $x$ and $y$ are identified by the ranking of approval probabilities in any lists in which $x$ and $y$ are in the same position (see example).

Relabel alternatives in decreasing order of preference: $x_1, x_2, \ldots, x_n$. 
Proof (sketch)

Preferences between any two alternatives $x$ and $y$ are identified by the ranking of approval probabilities in any lists in which $x$ and $y$ are in the same position (see example).

Relabel alternatives in decreasing order of preference: $x_1, x_2, ..., x_n$.

The approval prob of the worst alternative $x_n$ in a list where it is in last position identifies $\pi(n, x_n)$.

- If $x_n$ is moved up one position, then:
  - $x_n$ still chosen when the threshold is $x_n$ and the capacity is maximal;
  - but now also chosen with the same threshold when capacity is only $n-1$.

Hence the difference in approval when $x_n$ is in last position and when it is in position $n-1$ identifies $\pi(n-1, x_n)$:

Pushing $x_n$ up in the list notch by notch then pins down $\pi(n-2, x_n), ..., \pi(1, x_n)$. 
Preferences between any two alternatives $x$ and $y$ are identified by the ranking of approval probabilities in any lists in which $x$ and $y$ are in the same position (see example).

Relabel alternatives in decreasing order of preference: $x_1, x_2, \ldots, x_n$.

The approval prob of the worst alternative $x_n$ in a list where it is in last position identifies $\pi(n, x_n)$.

If $x_n$ is moved up one position, then:
- $x_n$ still chosen when the threshold is $x_n$ and the capacity is maximal;
- but now also chosen with the same threshold when capacity is only $n - 1$.

Hence the difference in approval when $x_n$ is in last position and when it is in position $n - 1$ identifies $\pi(n - 1, x_n)$:
Proof (sketch)

Preferences between any two alternatives \( x \) and \( y \) are identified by the ranking of approval probabilities in any lists in which \( x \) and \( y \) are in the same position (see example).

Relabel alternatives in decreasing order of preference: \( x_1, x_2, \ldots, x_n \).

The approval prob of the worst alternative \( x_n \) in a list where it is in last position identifies \( \pi(n, x_n) \).

If \( x_n \) is moved up one position, then:
- \( x_n \) still chosen when the threshold is \( x_n \) and the capacity is maximal;
- but now also chosen with the same threshold when capacity is only \( n - 1 \).

Hence the difference in approval when \( x_n \) is in last position and when it is in position \( n - 1 \) identifies \( \pi(n - 1, x_n) \):

Pushing \( x_n \) up in the list notch by notch then pins down
\( \pi(n - 2, x_n), \ldots, \pi(1, x_n) \).
Consider the difference in approval between the $k^{th}$ best alternative and the $(k + 1)^{th}$ best alternative when they are last.

The only event in which $x_k$ is approved while $x_{k+1}$ is not is when the threshold is $x_k$ and the capacity is $n$ (if capacity $< n$ or threshold $< x_k$ then neither is approved).

Hence the difference pins down $\pi(n, x_k)$ for all $k < n$. 
Fix a position $j < n$ and for any $k < n$ compare the approval probs of $x_k$ and $x_{k+1}$ at position $j$.

The difference in approval probs is the prob of all the events in which the threshold is $x_k$ and the capacity is at least $j$. Namely
\[ \pi(j, x_k) + \ldots + \pi(n, x_k) \]
Fix a position $j < n$ and for any $k < n$ compare the approval probs of $x_k$ and $x_{k+1}$ at position $j$.

The difference in approval probs is the prob of all the events in which the threshold is $x_k$ and the capacity is at least $j$. Namely

$$\pi (j, x_k) + ... + \pi (n, x_k)$$

Repeat the exercise with position $j + 1$: now the difference is

$$\pi (j + 1, x_k) + ... + \pi (n, x_k)$$
Fix a position $j < n$ and for any $k < n$ compare the approval probs of $x_k$ and $x_{k+1}$ at position $j$.

The difference in approval probs is the prob of all the events in which the threshold is $x_k$ and the capacity is at least $j$. Namely

$$\pi(j, x_k) + \ldots + \pi(n, x_k)$$

Repeat the exercise with position $j + 1$: now the difference is

$$\pi(j + 1, x_k) + \ldots + \pi(n, x_k)$$

As good applied economists, now take a diff-in-diff to identify $\pi(j, x_k)$. 
In ACM, we need an additional assumption for full identification.
In ACM, we need an additional assumption for full identification.

**Theorem**

In the ACM, preferences are uniquely identified. Moreover, if the probability distributions on thresholds and capacities are independent, they are uniquely identified by approval probabilities.
In ACM, we need an additional assumption for full identification.

**Theorem**

*In the ACM, preferences are uniquely identified. Moreover, if the probability distributions on thresholds and capacities are independent, they are uniquely identified by approval probabilities.*

The need for some restriction is seen from the simplest example: $x \succ y$. Preferences are identified as in DCM. However...
...although we also identify $\pi(2, y) = p(y, xy)$ and $\pi(1, y) = p(y, yx) - p(y, xy)$

it is impossible to break down $\pi(1, x) + \pi(2, x)$ from $\pi(1, x) + \pi(2, x) + \pi(2, y) = p(x, yx)$ and $1 = p(x, xy)$.

<table>
<thead>
<tr>
<th>ACM</th>
<th>$p(x, \lambda)$</th>
<th>$p(y, \lambda)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = xy$</td>
<td>$\pi_{1,x} + \pi_{2,x} + \pi_{1,y} + \pi_{2,y}$</td>
<td>$\pi_{2,y}$</td>
</tr>
<tr>
<td>$\lambda = yx$</td>
<td>$\pi_{1,x} + \pi_{2,x} + \pi_{2,y}$</td>
<td>$\pi_{1,y} + \pi_{2,y}$</td>
</tr>
</tbody>
</table>
In general, for any $k < n$, we can only hope to identify $\sum_{l \geq k} \pi(l, x_k)$. On the other hand, assume independence, i.e. $\pi(i, t) = \pi_i \pi_t$. Then in the two alternative case we have:

$$\pi_y = p(y, x)$$

identifying $\pi_y$ (and therefore $\pi_x$)

$$\pi_2 = p(y, x)$$

identifying $\pi_2$ (and therefore $\pi_1$)

The fact that $\pi_1$ is identified by the approval probs of only $x_2$ does generalise: the approval probs of $x_{k+1}, \ldots, x_n$ identify $\pi_k$.

This is the key for the recursive identifying algorithm in the proof (spared).
In general, for any $k < n$, we can only hope to identify $\sum_{l \geq k} \pi(l, x_k)$.

On the other hand, assume independence, i.e. $\pi(i, t) = \pi_i \pi_t$. Then in the two alternative case we have:

$$
\pi(y) = p(y, x) \text{ identifying } \pi(y)
$$

$$
\pi(y) \pi_2 = p(y, x) \text{ identifying } \pi_2 \text{ and therefore } \pi_1
$$

The fact that $\pi_1$ is identified by the approval probs of only $x_2$, does generalise: the approval probs of $x_{k+1}, \ldots, x_n$ identify $\pi_k$. This is the key for the recursive identifying algorithm in the proof (spared).
In general, for any $k < n$, we can only hope to identify $\sum_{l \geq k} \pi(l, x_k)$.

On the other hand, assume independence, i.e. $\pi(i, t) = \pi_i \pi_t$. Then in the two alternative case we have:

$\pi_y = p(y, yx)$ identifying $\pi_y$ (and therefore $\pi_x$)

The fact that $\pi_1$ is identified by the approval probs of only $x_2$ does generalise: the approval probs of $x_{k+1}, \ldots, x_n$ identify $\pi_k$.

This is the key for the recursive identifying algorithm in the proof (spared).
In general, for any $k < n$, we can only hope to identify $\sum_{l \geq k} \pi(l, x_k)$.

On the other hand, assume independence, i.e. $\pi(i, t) = \pi_i \pi_t$. Then in the two alternative case we have:

- $\pi_y = p(y, yx)$ identifying $\pi_y$ (and therefore $\pi_x$)
- $\pi_y \pi_2 = p(y, xy)$, identifying $\pi_2$ (and therefore $\pi_1$)
In general, for any $k < n$, we can only hope to identify $\sum_{l \geq k} \pi(l, x_k)$.

On the other hand, assume independence, i.e. $\pi(i, t) = \pi_i \pi_t$. Then in the two alternative case we have:

$\pi_y = p(y, yx)$ identifying $\pi_y$ (and therefore $\pi_x$)

$\pi_y \pi_2 = p(y, xy)$, identifying $\pi_2$ (and therefore $\pi_1$)

The fact that $\pi_1$ is identified by the approval probs of only $x_2$ does generalise: the approval probs of $x_{k+1}, \ldots, x_n$ identify $\pi_k$.

This is the key for the recursive identifying algorithm in the proof (spared).
Primitives can be learned from the observation of approval behaviour across lists.
Primitives can be learned from the observation of approval behaviour across lists.

Suppose now you:
(1) can control the lists, and
(2) have an objective you want to maximise (list is a choice variable).
Primitives can be learned from the observation of approval behaviour across lists.

Suppose now you:
(1) can control the lists, and
(2) have an objective you want to maximise (list is a choice variable).

Which list maximises the objective?
Suppose you want to max the **total number of approvals** (with the ’large number’ assumption that approval probs are identified with the fraction of times, over a large total of times, that an alternative is approved).

This objective makes sense in several instances:
- maximise the number of clicks;
- maximise the number of news pieces read;
- maximise social network involvement through Likes and sharing;
- maximise the size of an online shopping cart, etc.

We consider a (much) more general objective: maximize a weighted sum $\sum p(x, \lambda)$. The weights $w(x)$ allow to include objectives such as revenue per click or favouring some specific alternatives.
Suppose you want to max the **total number of approvals** (with the 'large number' assumption that approval probs are identified with the fraction of times, over a large total of times, that an alternative is approved).

This objective makes sense in several instances:
- maximise the number of clicks;
- maximise the number of news pieces read;
- maximise social network involvement through Likes and sharing;
- maximise the size of an online shopping cart, etc.
Suppose you want to max the **total number of approvals** (with the 'large number' assumption that approval probs are identified with the fraction of times, over a large total of times, that an alternative is approved).

This objective makes sense in several instances:
- maximise the number of clicks;
- maximise the number of news pieces read;
- maximise social network involvement through Likes and sharing;
- maximise the size of an online shopping cart, etc.

We consider a (much) more general objective: maximize a **weighted sum** of the $p(x, \lambda)$. The weights $w(x)$ allow to include objectives such as revenue per click or favouring some specific alternatives.
Two benchmark results in list design

The two models have very contrasting implications for list design in respect of the stated objective.
Two benchmark results in list design

The two models have very contrasting implications for list design in respect of the stated objective.

**Theorem**

*In the ACM, a list $\lambda$ is optimal iff it agrees with order of the weights $w(x)$, i.e. $w(x) > w(y) \Rightarrow x \lambda y$.*

**Corollary** (*List Invariance Principle*): In the ACM, if the weights are all the same, then any list is optimal.
Two benchmark results in list design

The two models have very contrasting implications for list design in respect of the stated objective.

**Theorem**

In the ACM, a list $\lambda$ is optimal iff it agrees with order of the weights $w(x)$, i.e. $w(x) > w(y) \Rightarrow x \lambda y$.

**Corollary** (*List Invariance Principle*): In the ACM, if the weights are all the same, then any list is optimal.

**Flash proof of Corollary** (Credit: Yuhta Ishii): Take any capacity-threshold pair $(i, t)$.

1) If $|\{x : x \succeq t\}| = k \leq i$, then $k$ items are approved.
2) Otherwise, $i$ items are approved.
3) Neither $k$ nor $i$ depends on the list. QED
Theorem

In the DCM:
1) If all weights are the same, then the unique maximiser of the number of approvals is the list that coincides with the preference order.
2) If the weights can differ and the probability distributions on thresholds and capacities are independent, then a list \( \lambda \) is optimal iff

\[
    w(x) \sum_{x \sim t} \pi(t) > w(y) \sum_{y \sim t} \pi(t) \Rightarrow x \lambda y
\]

3) In general, a list is optimal iff...
Since there are finitely many lists, a maximiser exists.

Suppose $x \succ y$ and look at swaps.
1. Take a list $\lambda$ in which $\lambda(x) > \lambda(y)$
2. Swap $x$ and $y$. Let $\lambda'$ be the same as $\lambda$ apart from $\lambda'(y) = \lambda(x)$ and $\lambda'(x) = \lambda(y)$
Intuition for proof with equal weights (DCM)

Since there are finitely many lists, a maximiser exists.

Suppose $x \succ y$ and look at swaps.
1. Take a list $\lambda$ in which $\lambda(x) > \lambda(y)$
2. Swap $x$ and $y$. Let $\lambda'$ be the same as $\lambda$ apart from $\lambda'(y) = \lambda(x)$ and $\lambda'(x) = \lambda(y)$

The approval probs of all $z$ different from $x$ and $y$ are not affected by the swap.
Intuition for proof with equal weights (DCM)

Since there are finitely many lists, a maximiser exists.

Suppose $x \succ y$ and look at swaps.
1. Take a list $\lambda$ in which $\lambda(x) > \lambda(y)$
2. Swap $x$ and $y$. Let $\lambda'$ be the same as $\lambda$ apart from $\lambda'(y) = \lambda(x)$ and $\lambda'(x) = \lambda(y)$

The approval probs of all $z$ different from $x$ and $y$ are not affected by the swap.

Any loss for $y$ is a gain for $x$. If $(i, t)$ leads to the approval of $y$ in $\lambda$ but not in $\lambda'$, then $y \succeq t$ and $i = \lambda(y)$. Hence $(i, t)$ leads to the approval of $x$ in $\lambda'$ and not in $\lambda$. 

Manzini Mariotti Ülkü  
Sequential Approval  
BRIColumbia 25 / 36
Since there are finitely many lists, a maximiser exists.

Suppose \( x \succ y \) and look at swaps.
1. Take a list \( \lambda \) in which \( \lambda(x) > \lambda(y) \)
2. Swap \( x \) and \( y \). Let \( \lambda' \) be the same as \( \lambda \) apart from \( \lambda'(y) = \lambda(x) \) and \( \lambda'(x) = \lambda(y) \)

The approval probs of all \( z \) different from \( x \) and \( y \) are not affected by the swap.

Any loss for \( y \) is a gain for \( x \). If \((i, t)\) leads to the approval of \( y \) in \( \lambda \) but not in \( \lambda' \), then \( y \succeq t \) and \( i = \lambda(y) \). Hence \((i, t)\) leads to the approval of \( x \) in \( \lambda' \) and not in \( \lambda \).

Some gain for \( x \) is not a loss for \( y \). The pair \((i, t) = (\lambda(y), x)\) leads to the approval of \( x \) in \( \lambda' \) but not in \( \lambda \). But it never leads to the approval of \( y \).
Intuition for proof (ACM)

As in DCM, a maximiser exists.

Suppose \( x \succ y \).

1. Take a list \( \lambda \) in which \( \lambda(x) = \lambda(y) + 1 \)
2. Swap to \( \lambda' \) - the same as \( \lambda \) apart from \( \lambda'(y) = \lambda(x) \) and \( \lambda'(x) = \lambda(y) \)

Because \( x \) and \( y \) are adjacent, this can only affect the approval probabilities of \( x \) and \( y \).

Any loss for \( y \) is a gain for \( x \). Suppose \( (i, t) \) leads to the approval of \( y \) in \( \lambda \) but not in \( \lambda' \). Then \( (i, t) \) cannot lead to the approval of \( x \) in \( \lambda \) (capacity is exhausted) but has to lead to the approval of \( x \) in \( \lambda' \).
As in DCM, a maximiser exists.

Suppose $x \succ y$.

1. Take a list $\lambda$ in which $\lambda (x) = \lambda (y) + 1$
2. Swap to $\lambda'$ - the same as $\lambda$ apart from $\lambda' (y) = \lambda (x)$ and $\lambda' (x) = \lambda (y)$

Because $x$ and $y$ are adjacent, this can only affect the approval probabilities of $x$ and $y$. 
As in DCM, a maximiser exists.

Suppose $x \succ y$.
1. Take a list $\lambda$ in which $\lambda(x) = \lambda(y) + 1$
2. Swap to $\lambda'$ - the same as $\lambda$ apart from $\lambda'(y) = \lambda(x)$ and $\lambda'(x) = \lambda(y)$

Because $x$ and $y$ are adjacent, this can only affect the approval probabilities of $x$ and $y$.

*Any loss for $y$ is a gain for $x$. Suppose $(i, t)$ leads to the approval of $y$ in $\lambda$ but not in $\lambda'$. Then $(i, t)$ cannot lead to the approval of $x$ in $\lambda$ (capacity is exhausted) but has to lead to the approval of $x$ in $\lambda'$.**
Is any gain for \( x \) a loss for \( y \)?
Intuition for proof (ACM) - ctnd.

Is any gain for $x$ a loss for $y$?

Yes - *this is the crucial difference* from DCM.
Is any gain for $x$ a loss for $y$?

Yes - *this is the crucial difference* from DCM.

If $(i, t)$ leads to the approval of $x$ in $\lambda'$ but not in $\lambda$, then $y$ must have exhausted capacity in $\lambda$, meaning $y \preceq t$. Furthermore capacity cannot have been exhausted in $\lambda$ when the agent reaches $y$. Hence $(i, t)$ must lead to the approval of $y$ in $\lambda$. But since $x$ will consume the last bit of capacity, $(i, t)$ cannot lead to the approval of $y$ in $\lambda'$. 
Is any gain for $x$ a loss for $y$?

Yes - *this is the crucial difference* from DCM.

If $(i, t)$ leads to the approval of $x$ in $\lambda'$ but not in $\lambda$, then $y$ must have exhausted capacity in $\lambda$, meaning $y \preceq t$. Furthermore capacity cannot have been exhausted in $\lambda$ when the agent reaches $y$. Hence $(i, t)$ must lead to the approval of $y$ in $\lambda$. But since $x$ will consume the last bit of capacity, $(i, t)$ cannot lead to the approval of $y$ in $\lambda'$.

(In DCM $x$ gains from the swap in events that do not benefit $y$ before the swap. In ACM this cannot happen: $y$ must have absorbed capacity pre-swap for $x$ to gain from the swap. Hence $y$ loses what $x$ gains.)
Comparatives

For a given preference $\succeq$ consider probability distributions $\pi_a$ and $\pi_b$ and their associated approval functions $p_a$ and $p_b$, respectively.

Say that $a$ is *strongly more approving* than $b$ iff $p_a(x, \lambda) \geq p_b(x, \lambda)$ for all alternatives $x$ and lists $\lambda$.

Manzini Mariotti Ülkü
Sequential Approval
BRIColumbia
Comparatives

For a given preference $\succeq$ consider probability distributions $\pi_a$ and $\pi_b$ and their associated approval functions $p_a$ and $p_b$, respectively.

Say that $a$ is *strongly more approving* than $b$ iff $p_a(x, \lambda) \geq p_b(x, \lambda)$ for all alternatives $x$ and lists $\lambda$.

Say that $a$ is *weakly more approving* than $b$ iff, for any list $\lambda$, the total number of approvals by $a$ in $\lambda$ is greater than that by $b$, i.e. $\sum_x p_a(x, \lambda) \geq \sum_x p_b(x, \lambda)$.
Comparatives

For a given preference $\succeq$ consider probability distributions $\pi_a$ and $\pi_b$ and their associated approval functions $p_a$ and $p_b$, respectively.

Say that $a$ is strongly more approving than $b$ iff $p_a(x, \lambda) \geq p_b(x, \lambda)$ for all alternatives $x$ and lists $\lambda$.

Say that $a$ is weakly more approving than $b$ iff, for any list $\lambda$, the total number of approvals by $a$ in $\lambda$ is greater than that by $b$, i.e.
\[ \sum_x p_a(x, \lambda) \geq \sum_x p_b(x, \lambda) . \]

Numbering the alternatives from best to worst, any $\pi$ defines uniquely a (univariate) numerical random variable $X_{\pi}$ on $\{1, \ldots, n\}$ that gives the minimum of any capacity-threshold pair $(m, i)$, i.e.
\[ \Pr (X_{\pi} = i) = \pi (\{(m, x_j) : \min (m, j) = i\}) \]
Given a $\pi_a$, let $F_a$ denote the cdf.

**Theorem.** In the DCM, $a$ is strongly more approving than $b$ if and only if $F_a$ first order stochastically dominates $F_b$. 
Given a $\pi_a$, let $F_a$ denote the cdf.

**Theorem.** In the DCM, $a$ is strongly more approving than $b$ if and only if $F_a$ first order stochastically dominates $F_b$.

**Theorem.** In the ACM, $a$ is weakly more approving than $b$ if and only if

$$\mathbb{E}(X_{\pi_a}^-) \geq \mathbb{E}(X_{\pi_b}^-).$$
How do we know whether the agent is representable through DCM or ACM?

A1. If $\lambda(x) = \lambda'(x)$, then $p(x,\lambda) = p(x,\lambda')$ (only the position matters).

A2. If $\lambda(x) < \lambda'(x)$, then $p(x,\lambda) > p(x,\lambda')$. (higher positions are better)

A3. If $\lambda(x) = \lambda'(y) = k$, $\mu(x) = \mu'(y) = k - 1$ and $p(x,\lambda) > p(y,\lambda')$ then $p(x,\mu) - p(y,\mu') > p(x,\lambda) - p(y,\lambda')$. (supermodularity in quality and position)

A4a. There exists $x$ such that if $\lambda(x) = 1$, then $p(x,\lambda) = 1$. (dominant alternative)

A4b. For all $x$ and $\lambda$, $p(x,\lambda) > 0$. (positivity)

A4c. If $p(x,\lambda) = p(y,\lambda')$ and $\lambda(x) = \lambda'(y)$ then $x = y$. (linearity)
How do we know whether the agent is representable through DCM or ACM?

A1. If $\lambda(x) = \lambda'(x)$, then $p(x, \lambda) = p(x, \lambda')$ (only the position matters)
How do we know whether the agent is representable through DCM or ACM?

A1. If $\lambda(x) = \lambda'(x)$, then $p(x, \lambda) = p(x, \lambda')$ (only the position matters)

A2. If $\lambda(x) < \lambda'(x)$, then $p(x, \lambda) > p(x, \lambda')$. (higher positions are better)
How do we know whether the agent is representable through DCM or ACM?

A1. If $\lambda(x) = \lambda'(x)$, then $p(x, \lambda) = p(x, \lambda')$ (only the position matters)

A2. If $\lambda(x) < \lambda'(x)$, then $p(x, \lambda) > p(x, \lambda')$. (higher positions are better)

A3. If $\lambda(x) = \lambda'(y) = k$, $\mu(x) = \mu'(y) = k - 1$ and $p(x, \lambda) > p(y, \lambda')$ then

$$p(x, \mu) - p(y, \mu') > p(x, \lambda) - p(y, \lambda').$$

(supermodularity in quality and position)
How do we know whether the agent is representable through DCM or ACM?

A1. If $\lambda(x) = \lambda'(x)$, then $p(x, \lambda) = p(x, \lambda')$ (only the position matters)

A2. If $\lambda(x) < \lambda'(x)$, then $p(x, \lambda) > p(x, \lambda')$. (higher positions are better)

A3. If $\lambda(x) = \lambda'(y) = k$, $\mu(x) = \mu'(y) = k - 1$ and $p(x, \lambda) > p(y, \lambda')$ then

$$p(x, \mu) - p(y, \mu') > p(x, \lambda) - p(y, \lambda').$$

(supermodularity in quality and position)

A4a. There exists $x$ such that if $\lambda(x) = 1$, then $p(x, \lambda) = 1$. (dominant alternative)

A4b. For all $x$ and $\lambda$, $p(x, \lambda) > 0$. (positivity)

A4c. If $p(x, \lambda) = p(y, \lambda')$ and $\lambda(x) = \lambda'(y)$ then $x = y$. (linearity)
Theorem. A stochastic approval function is a DCM if and only if it satisfies A1-A4.
Theorem. A stochastic approval function is a DCM if and only if it satisfies A1-A4.

Various extensions are possible with minor variations of the axioms:

- Preferences can be weak orders
- Depth can be zero
- Probabilities can be zero
B1. If $\lambda(x) \leq \lambda'(x)$, then $p(x, \lambda) \geq p(x, \lambda')$. (higher positions are weakly better)

B2. If $\lambda(x) = \lambda'(y) = k$, $\mu(x) = \mu'(y) = k - 1$ and $p(x, \lambda) \geq p(y, \lambda')$ then

$$p(x, \mu) - p(y, \mu') \geq p(x, \lambda) - p(y, \lambda').$$

(the advantage of better alternatives is weakly enhanced in higher positions)
B1. If $\lambda(x) \leq \lambda'(x)$, then $p(x, \lambda) \geq p(x, \lambda')$. (higher positions are weakly better)

B2. If $\lambda(x) = \lambda'(y) = k$, $\mu(x) = \mu'(y) = k - 1$ and $p(x, \lambda) \geq p(y, \lambda')$ then

$$p(x, \mu) - p(y, \mu') \geq p(x, \lambda) - p(y, \lambda').$$

(the advantage of better alternatives is weakly enhanced in higher positions)

Theorem

A stochastic approval function is a generalised DCM if and only if it satisfies B1 and B2.
Partly but not entirely parallel to that of DCM.

A1’ Only predecessor set matters
A2’ Smaller predecessor sets are better
A3’ Supermodularity in quality and smallness of predecessor set
A4’ Positivity, linearity, dominant alternative
A5’ If consecutive $x$ and $y$ are switched, $y$ gains exactly what $y$ loses.
Wishful Theorem. A stochastic approval function is an ACM only if it satisfies A1-A5’, and perhaps also if.

The big problem for a neat characterisation is that only the number of better alternatives counts in the predecessor set, whereas also the identity of worse alternatives counts.
Concluding remarks

A model of approval, not choice.

We have studied situations in which both the menu and the selections are typically 'large' (either 'pre-choice' or 'non-choice').
Concluding remarks

A model of approval, not choice.

We have studied situations in which both the menu and the selections are typically 'large' (either 'pre-choice' or 'non-choice').

Approval is intrinsically related to satisficing behaviour. We have provided two models that seems plausible, but others are possible.
Concluding remarks

A model of approval, not choice.

We have studied situations in which both the menu and the selections are typically 'large' (either 'pre-choice' or 'non-choice').

Approval is intrinsically related to satisficing behaviour. We have provided two models that seems plausible, but others are possible.

In particular, natural to look at non-stationary thresholds.
Concluding remarks

A model of approval, not choice.

We have studied situations in which both the menu and the selections are typically 'large' (either 'pre-choice' or 'non-choice').

Approval is intrinsically related to satisficing behaviour. We have provided two models that seems plausible, but others are possible.

In particular, natural to look at non-stationary thresholds.

List design seems to offer ample scope for further relevant research.
THANK YOU!