Edinburgh Napier UNIVERSITY

Strategy Choice in Transit Networks

Achille **Fonzone** – *Edinburgh Napier University*

Jan-Dirk **Schmöcker – ***Kyoto University*

Fumitaka **Kurauchi**, Seham M. **Hassan** – *Gifu University*

Strategies and hyperpaths

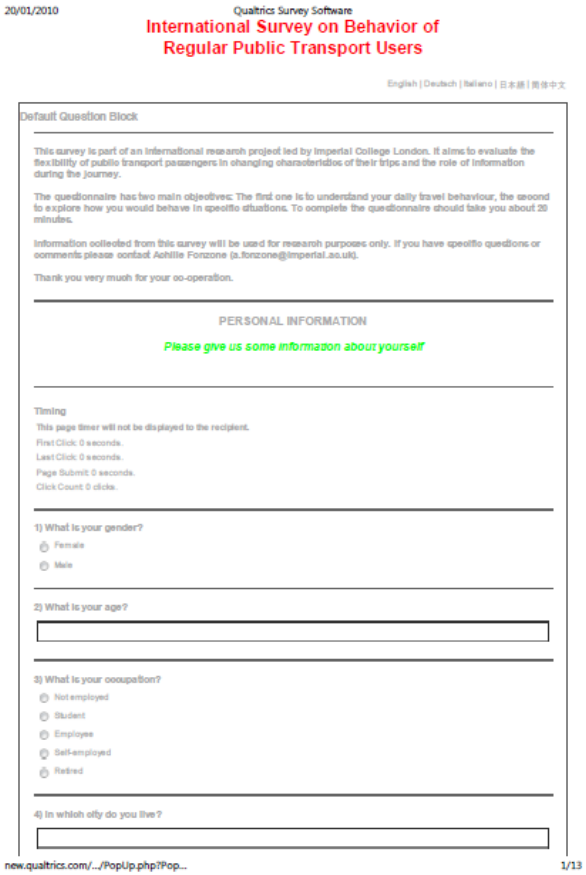
- Passengers may have the possibility of choosing more than one line to get (close) to his destination
  - Lines have different attractiveness in terms of travel time, number of changes, seat availability …
- They need a strategy: a set of attractive lines and a selection rule
  - Ignore “obviously” bad lines (Lampkin and Saalmans, 1967)
  - Minimise expected travel time assuming that the next vehicle is taken serving a line within the selected sub-set
- Strategic behaviour is the basis of the assignment of transit networks with high frequency services
  - Spiess and Florian (1989) combine strategies with equilibrium
  - Nguyen and Pallottino (1988) describe the problem in terms of hyperpaths

Do we really think in terms of hyperpaths?

- The behavioural model underpinning the hyperpath approach owes its appeal to the simplicity of its assumption
  - Expected trip time maximisers
  - Unbounded computing capacity
  - Perfect (stochastic as to line arrivals at stops) information
  - No real-time info available (apart from Gentile et al., 2005)
- But route choice is a complex spatial decision making process
  - Deriving from bounded rationality
  - Deeply affected by travel habits developed through implementation and evaluation of trial trips
  - Based on a wide range of criteria
- Can users think and act in terms of hyperpaths?
  - Does an attitude towards considering alternative trip solutions exist?
    - What is the effect of information?
  - Does the Spiess & Florian's model describe transit user behaviour correctly?

Survey

- Exploratory research
- Questionnaire
  - Personal information questions
  - Actual Behaviour
  - Hypothetical scenarios
- Web based, non random-sampling
  - Mainly transport students/scholars, engineering companies
- Targeting 6 countries: China, Germany, Japan, Italy, UK, and USA
- 579 complete replies
  - 1022 contacts


20/01/2010 Qualtrics Survey Software

International Survey on Behavior of Regular Public Transport Users

English | Deutsch | Italiano | 日本語 | 简体中文

Default Question Block

This survey is part of an international research project led by Imperial College London. It aims to evaluate the flexibility of public transport passengers in changing characteristics of their trips and the role of information during the journey.

The questionnaire has two main objectives: The first one is to understand your daily travel behaviour, the second to explore how you would behave in specific situations. To complete the questionnaire should take you about 20 minutes.

Information collected from this survey will be used for research purposes only. If you have specific questions or comments please contact Achille Fonzone (a.fonzone@imperial.ac.uk).

Thank you very much for your co-operation.

PERSONAL INFORMATION

Please give us some information about yourself

Timing

This page timer will not be displayed to the recipient.

First Click: 0 seconds.

Last Click: 0 seconds.

Page Submit: 0 seconds.

Click Count: 0 clicks.

1) What is your gender?

Female

Male

2) What is your age?

3) What is your occupation?

Not employed

Student

Employee

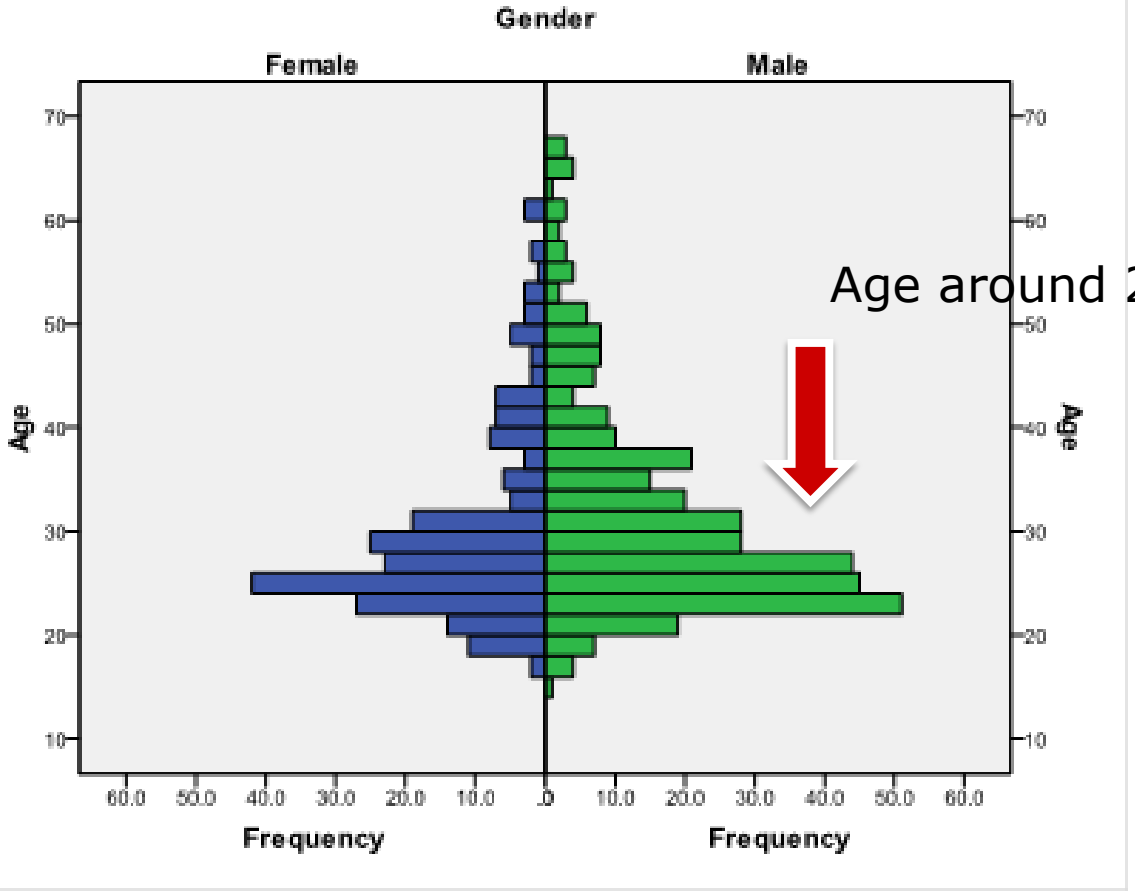
Self-employed

Retired

4) In which city do you live?

new.qualtrics.com/.../PopUp.php?Pop... 1/13

Sample

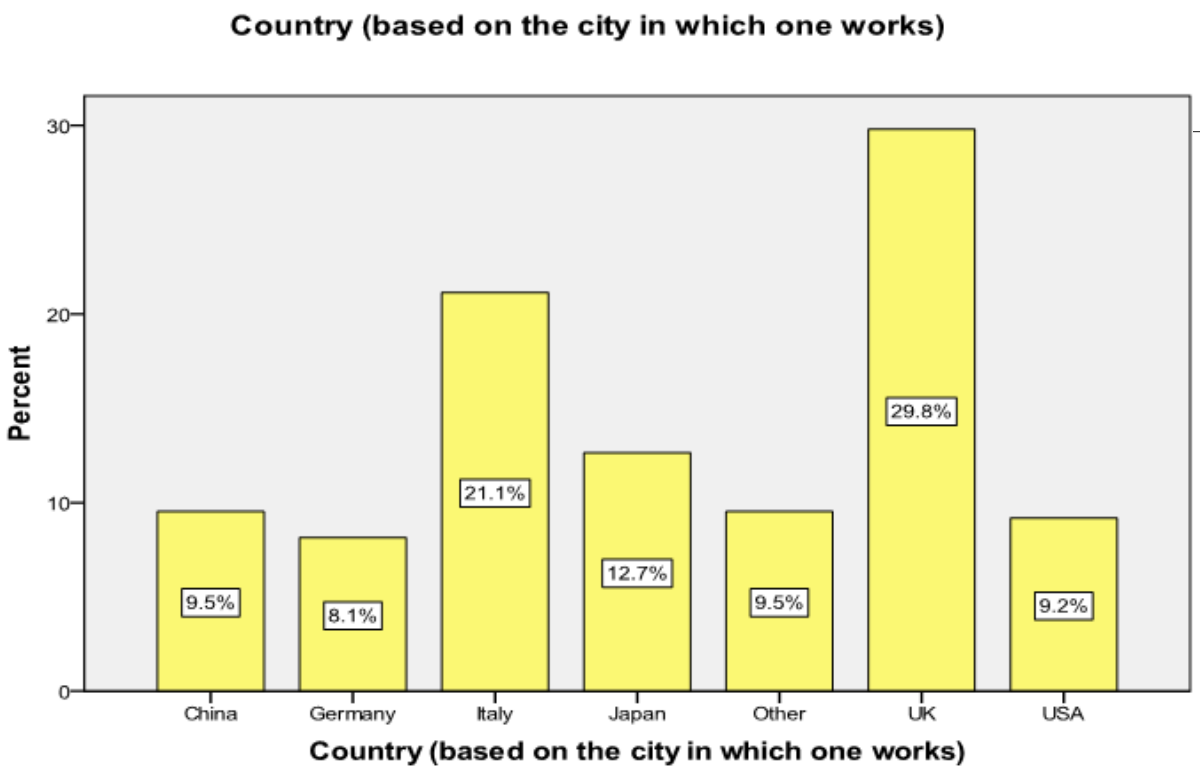


Age around 20-30

Sample


Occupation
Percent
50
40
30
20
10
0
48.3%
46.0%
1.7%
3.1%
0.9%
Not employed
Student
Employee
Self-employed
Retired
Occupation

Sample



City of job/study (first 10)

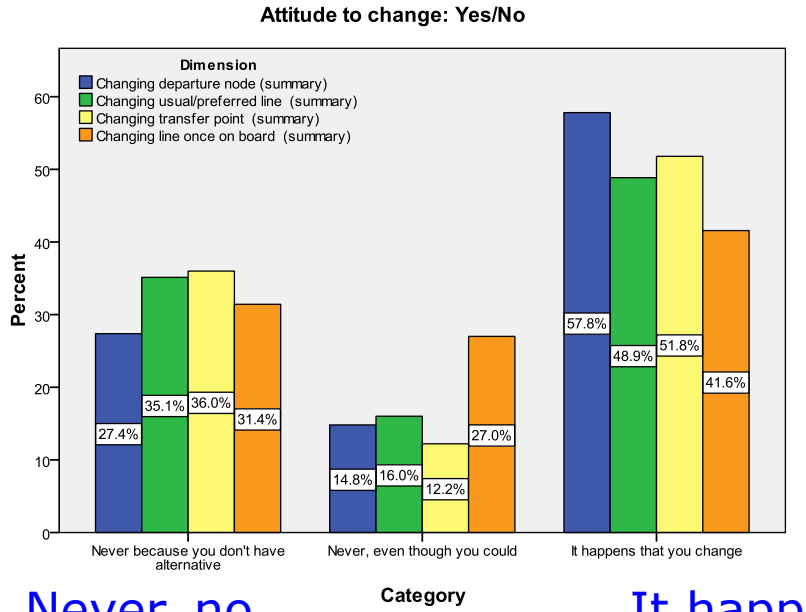
City	Country	Overall percent	Overall cumulative percent	Percent within the country
London	UK	24.9	24.9	79.8
Roma	Italy	13.4	38.3	60.7
Tokyo	Japan	7.4	45.7	54.1
Karlsruhe	Germany	4.9	50.5	58.7
Taranto	Italy	4.5	55.1	25.0
Wuhan	China	4.3	59.4	46.2
Berkeley	USA	2.5	61.9	25.0
Graz	Austria	2.3	64.3	81.3
Kyoto	Japan	2.3	66.6	17.6
New York	USA	2.0	68.6	19.6

Edinburgh Napier UNIVERSITY

Trip variations and information

Attitude towards changes

Q: how often does it happen that you decide to change dep point / line / transfer point from the usual one?

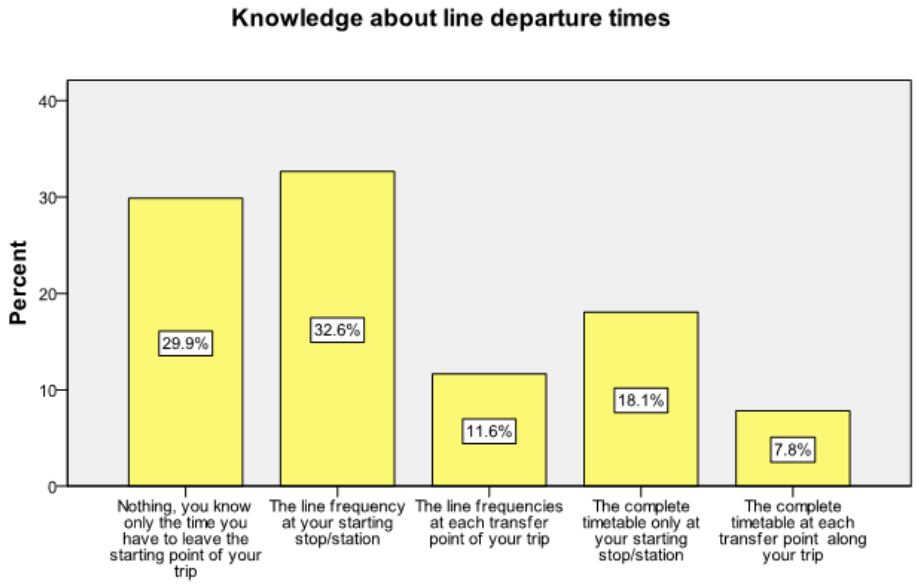


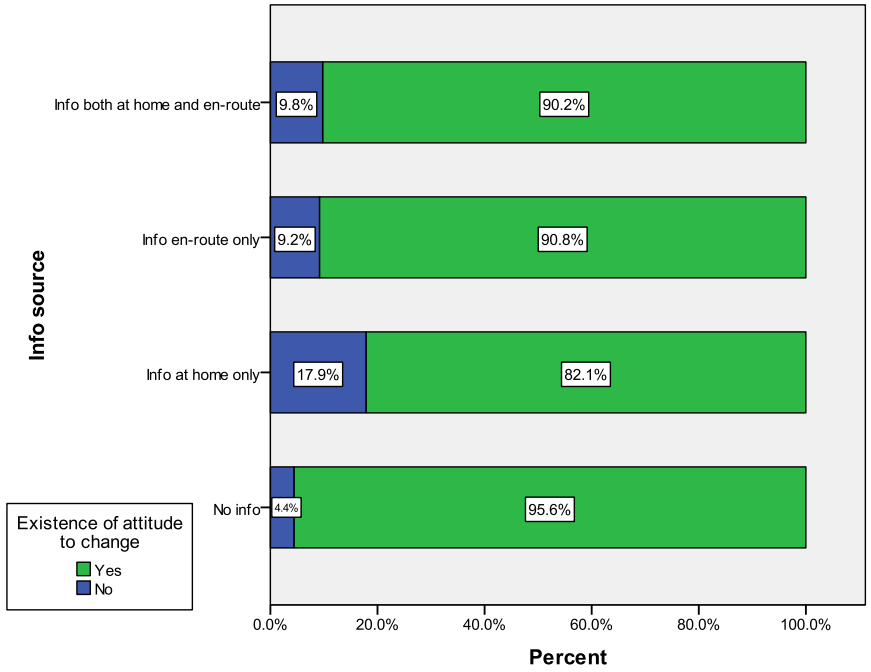
Never, no alternative

Never, even some alternatives

It happens

- Multiple attractive paths are considered supporting hyperpath-based models
- The most frequent kind of change concerns the departure stop/station
  - Whose choice is often ignored by models
- Transit network representations may be not consistent with travellers' mental maps





Knowledge and information

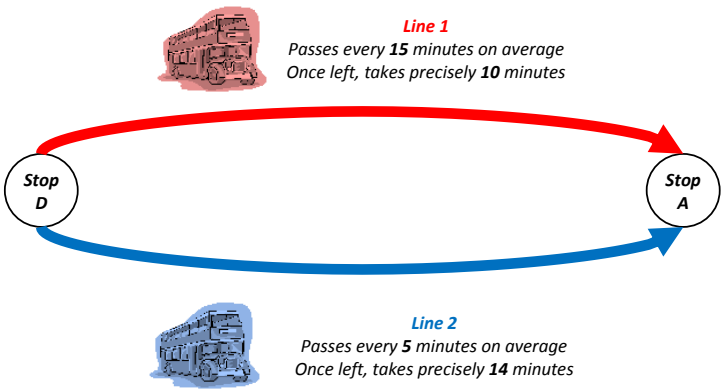
- Very few respondents have explicit knowledge about service timetables and frequencies at all the transfer points of their reported trips
  - Even though these are usual.
- A weaker attitude to change seems to be related to
  - A better knowledge on service departure times
  - The usage of external sources of information

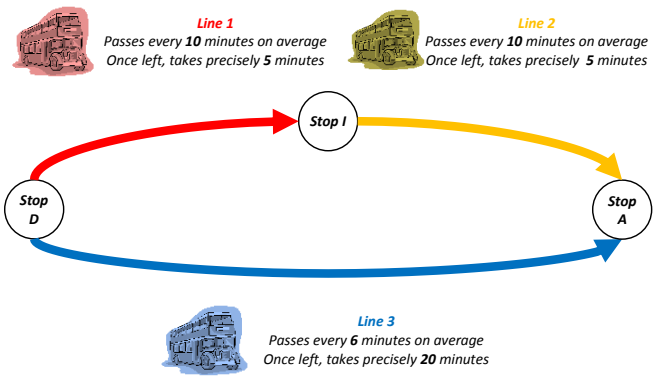
Edinburgh Napier UNIVERSITY

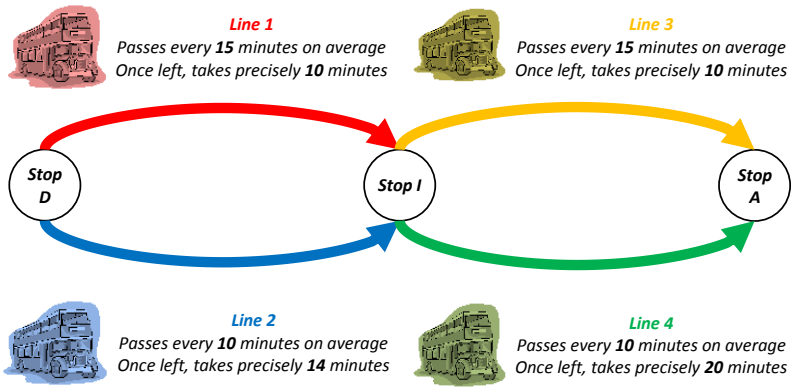
A cluster model of transit user stated behaviour

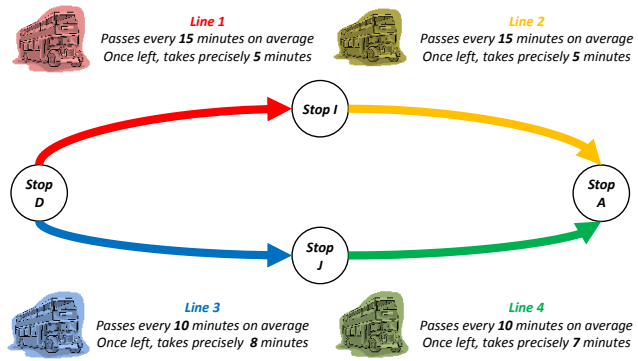
Validity of the S&F route choice assumption

SP experiments









Choice characteristics

	Line 1 (L1)				Line 2 (L2)				Line 1 + Line 2 (L1+L2)			
Scenario	On-board time	Waiting time	Total Travel Time	Number of Transfer	On-board time	Waiting time	Total Travel Time	Number of Transfer	On-board time	Waiting time	Total Travel Time	Number of Transfer
1 (a)	10	15	25	0	14	5	19	0	13	3.75	16.75	0
2 (b-1)	10	15	25	0	14	10	24	0	12.4	6	18.4	0
3 (b-2)	10	15	25	0	20	10	30	0	16	6	22	0
4 (d)	10	20	30	1	20	6	26	0	16.25	7.5	23.75	0.375
5 (d)	12	16	28	0	16	8	24	1	15.2	6.4	21.6	0.8
6 (e)	10	30	40	1	15	20	35	1	13	18	31	1

Estimation Result of Cross Nested Logit Model (Kurauchi et al., 2012)

User categories									
	Age	60+	60-						
	Country of residence	All	China	Others					
	Crowded train		All	Sometimes fail to board			others		
	Rate of travel time range			<75%	>75%		<75%		>75%
	Occupation			All	Student	others	Student	others	All
Num. of Observations		82	230	275	78	68	949	1109	521
Num. of Samples		14	39	47	13	13	164	190	90
Estimated parameters for each user category	Travel Time	-0.100 (-1.78)	**-0.154 (-4.71)****	**-0.395 (-8.84)****	-0.069 (-1.34)	**-0.518 (-5.05)****	**-0.302 (-12.64)****	**-0.319 (-12.64)****	**-0.336 (-10.56)****
	Waiting Time	**-0.214 (-5.92)****	**-0.183 (-9.40)****	**-0.282 (-10.80)****	**-0.142 (-4.86)****	**-0.382 (-6.26)****	**-0.212 (-15.25)****	**-0.254 (-15.11)****	**-0.258 (-12.99)****
	Number of Transfers	-0.447 (-0.79)	**-1.190 (-3.99)****	**-0.582 (-2.24)***	-0.195 (-0.41)	-0.476 (-0.78)	**-0.946 (-5.97)****	**-0.987 (-5.88)****	**-0.821 (-4.00)****
Estimated Variance for panel data		0.672(9.87)**							
Estimated Parameters for CNL model	Lambda1	**2.80(6.74)****							
	Lambda2	**1.58(7.76)****							
	alpha11	1.00(fixed)							
	alpha31	**0.510(7.37(=0), -7.07(=1))****							
	alpha22	1.00(fixed)							
	alpha32	**0.490(7.07(=0), -7.37(=1))****							
Number of observations		3312							
Number of samples		570							
Likelihood ratio test		2045.925							
Adjusted rho-square		0.273							

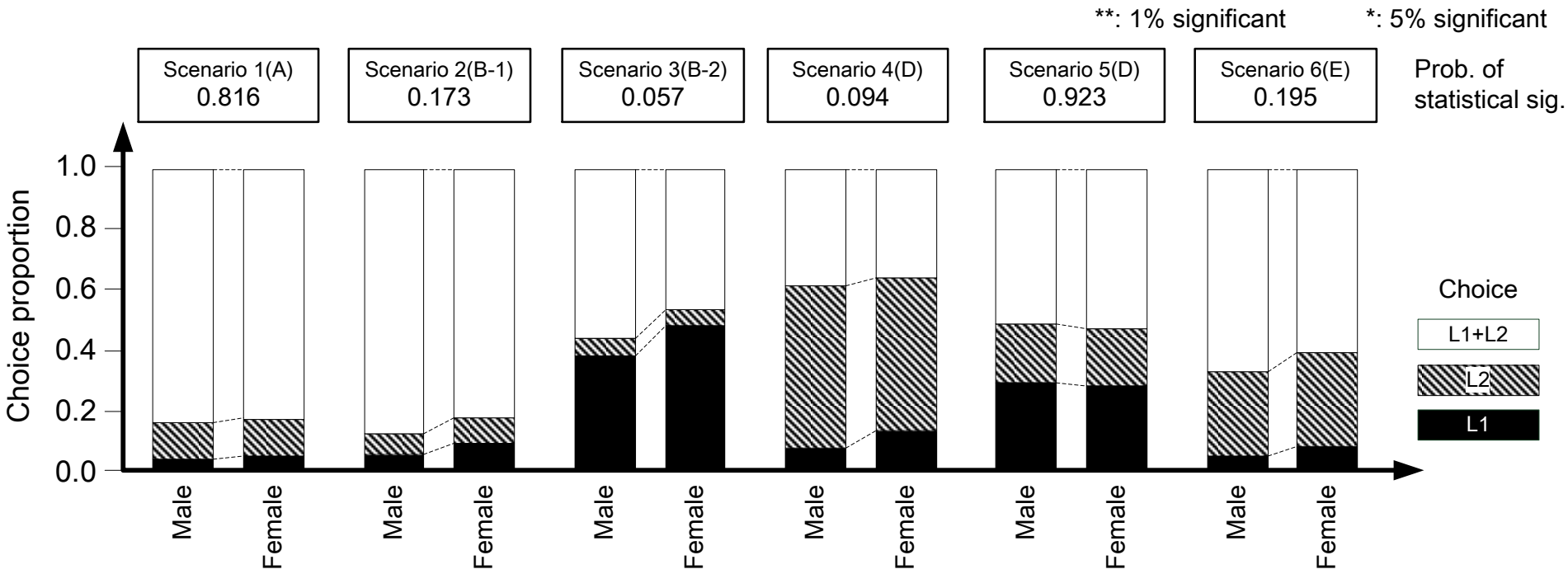
***: 5% significant, **: 1% significant**

Methodology: SPSS TwoSteps procedure

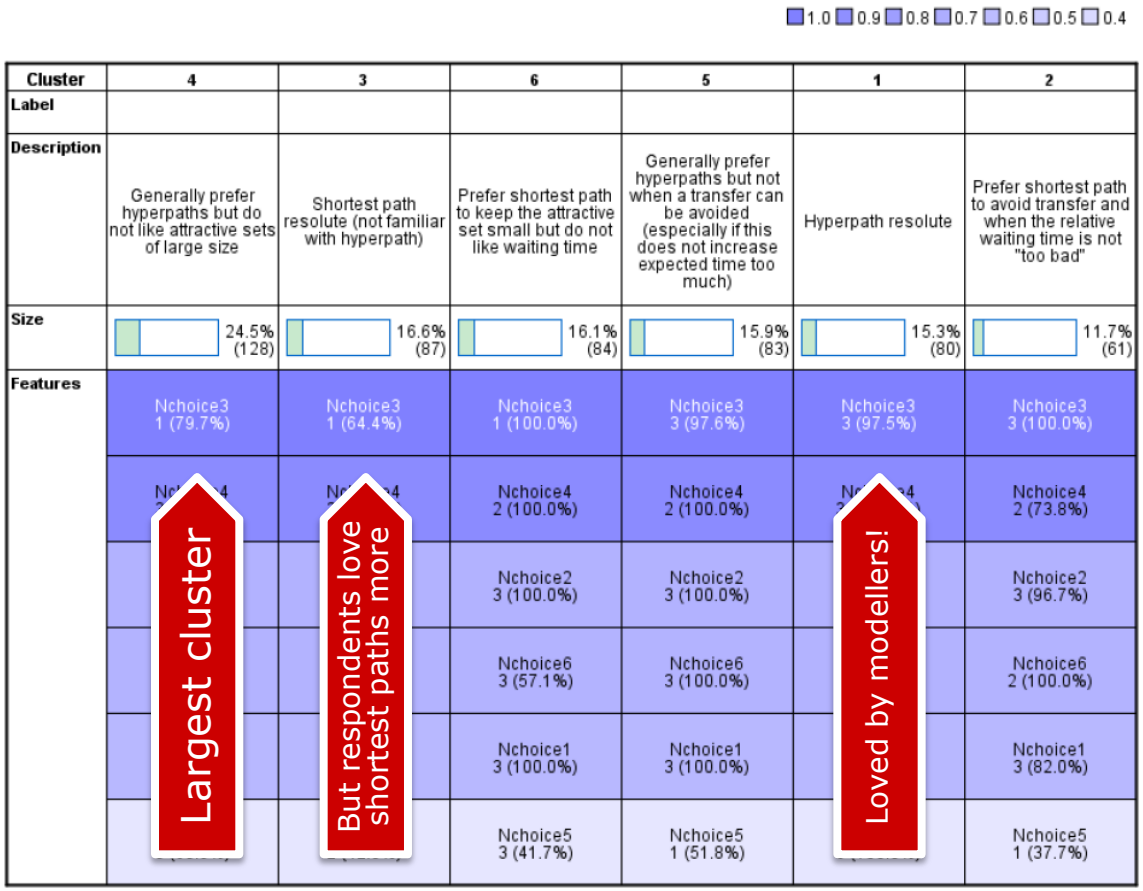
	Line 1 (L1)				Line 2 (L2)				Line 1 + Line 2 (L1+L2)			
Scenario	On-board time	Waiting time	Total Travel Time	Number of Transfer	On-board time	Waiting time	Total Travel Time	Number of Transfer	On-board time	Waiting time	Total Travel Time	Number of Transfer
1 (a)	10	15	25	0	14	5	19	0	13	3.75	16.75	0
2 (b-1)	10	15	25	0	14	10	24	0	12.4	6	18.4	0
3 (b-2)	10	15	25	0	20	10	30	0	16	6	22	0
4 (d)	10	20	30	1	20	6	26	0	16.25	7.5	23.75	0.375
5 (d)	12	16	28	0	16	8	24	1	15.2	6.4	21.6	0.8
6 (e)	10	30	40	1	15	20	35	1	13	18	31	1

- Clusterisation of behaviour based on SPSS TwoStep, an acknowledged standard when nominal variables are involved
  - Step 1 - Pre-clustering: following a sequential approach known as “cluster feature tree” a first clustering is performed
    - Each cluster is characterised by “number of records, mean and variance of each range field, and counts for each category of each symbolic field”
    - Pre-clusters are then used instead of original data in Step 2
    - The result of CF tree procedure depends on the input order of records
  - Step 2 – Clustering: A hierarchical clustering procedure is applied to preclusters
    - If the number of clusters is not fixed in advance, SPSS chooses the number of clusters based on BIC (or AIC)

Overall choice pattern



Behaviour clusters: Cluster features



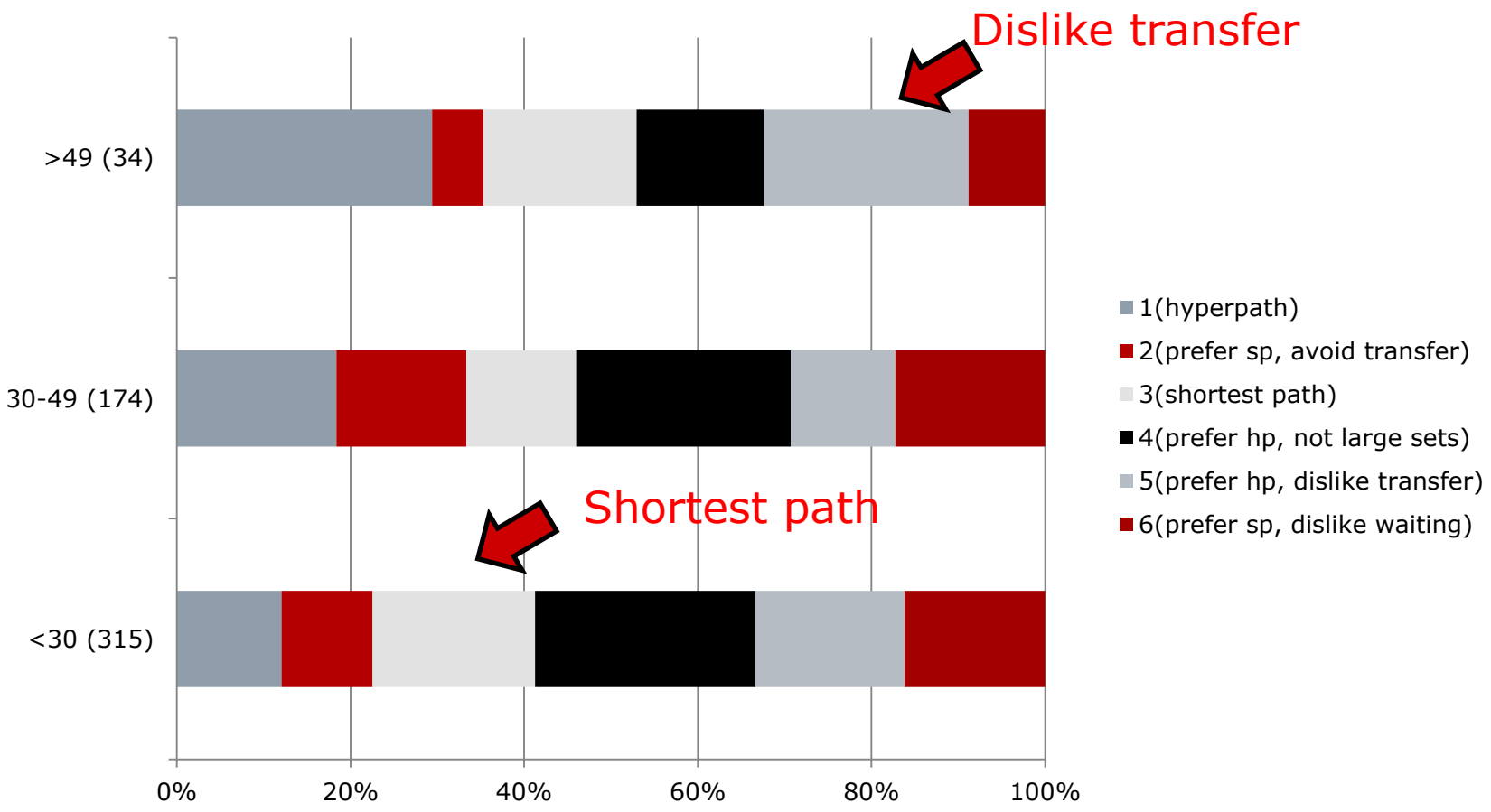
Cluster	4	3	6	5	1	2
Label						
Description	Generally prefer hyperpaths but do not like attractive sets of large size	Shortest path resolute (not familiar with hyperpath)	Prefer shortest path to keep the attractive set small but do not like waiting time	Generally prefer hyperpaths but not when a transfer can be avoided (especially if this does not increase expected time too much)	Hyperpath resolute	Prefer shortest path to avoid transfer and when the relative waiting time is not "too bad"
Size	24.5% (128)	16.6% (87)	16.1% (84)	15.9% (83)	15.3% (80)	11.7% (61)
Features	Nchoice3 1 (79.7%)	Nchoice3 1 (64.4%)	Nchoice3 1 (100.0%)	Nchoice3 3 (97.6%)	Nchoice3 3 (97.5%)	Nchoice3 3 (100.0%)
	[illegible]	[illegible]	Nchoice4 2 (100.0%)	Nchoice4 2 (100.0%)	[illegible]	Nchoice4 2 (73.8%)
			Nchoice2 3 (100.0%)	Nchoice2 3 (100.0%)		Nchoice2 3 (96.7%)
			Nchoice6 3 (57.1%)	Nchoice6 3 (100.0%)		Nchoice6 2 (100.0%)
			Nchoice1 3 (100.0%)	Nchoice1 3 (100.0%)		Nchoice1 3 (82.0%)
			Nchoice5 3 (41.7%)	Nchoice5 1 (51.8%)		Nchoice5 1 (37.7%)

Chi-square test on strategy choice

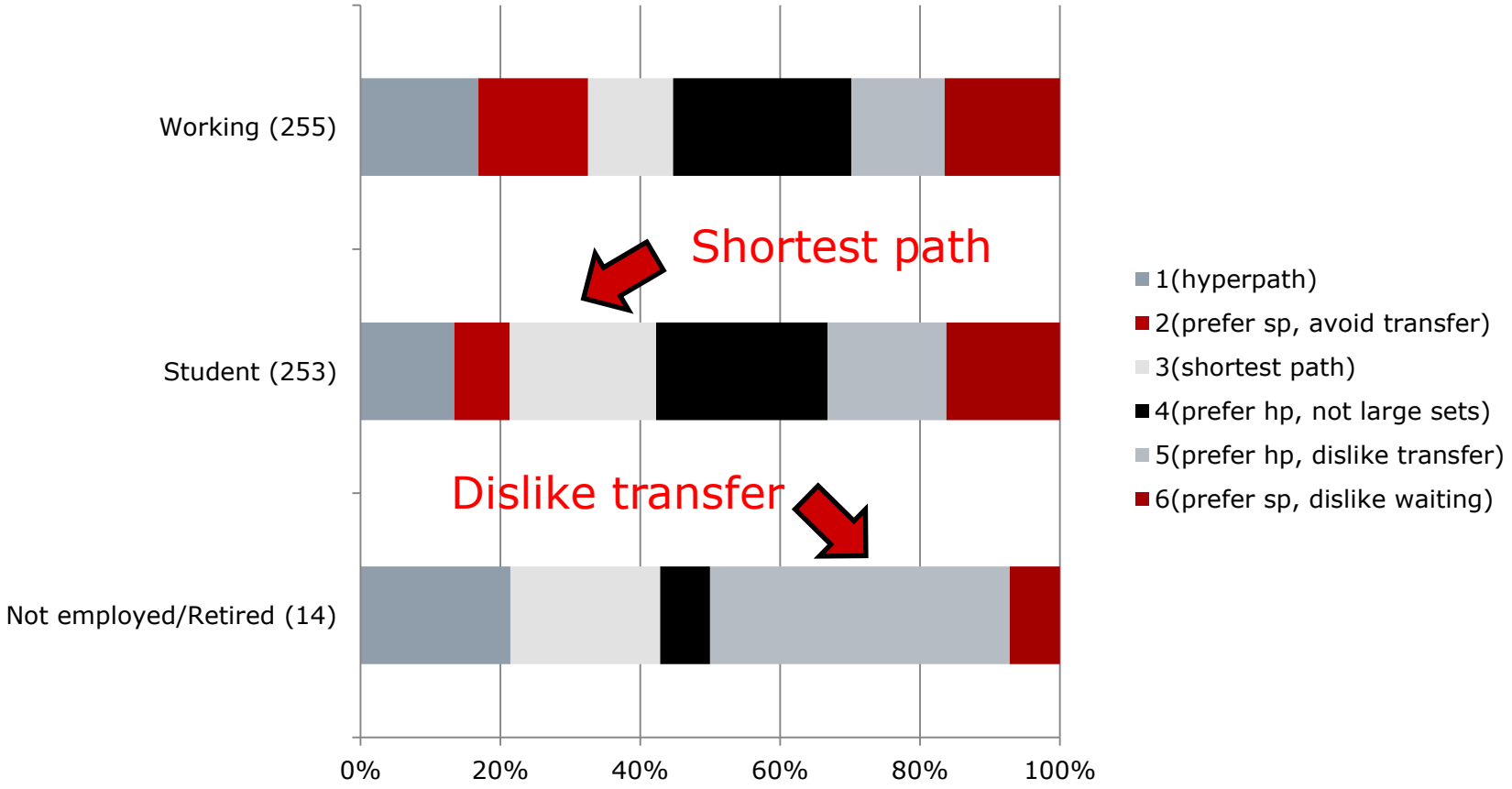
Behavioural choice	User/trip characteristic	Categories	Number of cases	Significance (asymptotic, 2-tails)
Behaviour cluster	Demand cluster	See Figure 3	394	0.44
	Gender	Male, Female	523	0.10
	Age	≤29-,30-49, ≥50	523	0.04
	Occupation	Student, Employee, Other	523	0.03
	Importance of punctuality	Not important to important: 1-2, 3, 4-5 of the original scale	520	0.84
	Travel time reliability*	0, [0;0.5), [0.5,1), ≥1	523	0.03
	Usual congestion	You can always find a seat, Sometimes you have to stand, You always have to stand, Sometimes you can't get onto the first vehicle	523	0.51
	Knowledge about service characteristics	[Regarding the departure times of the lines you use, the passenger knows] Only the departing time from the starting point of the trip, The line frequency at the starting stop/station, The line frequencies at each transfer point of the trip, The complete timetable only at the starting stop/station, The complete timetable at each transfer point along the trip	481	0.21

* $\frac{maxTT-minTT}{aveTT}$ where maxTT, minTT and aveTT are the maximum, the minimum, and the average travel time

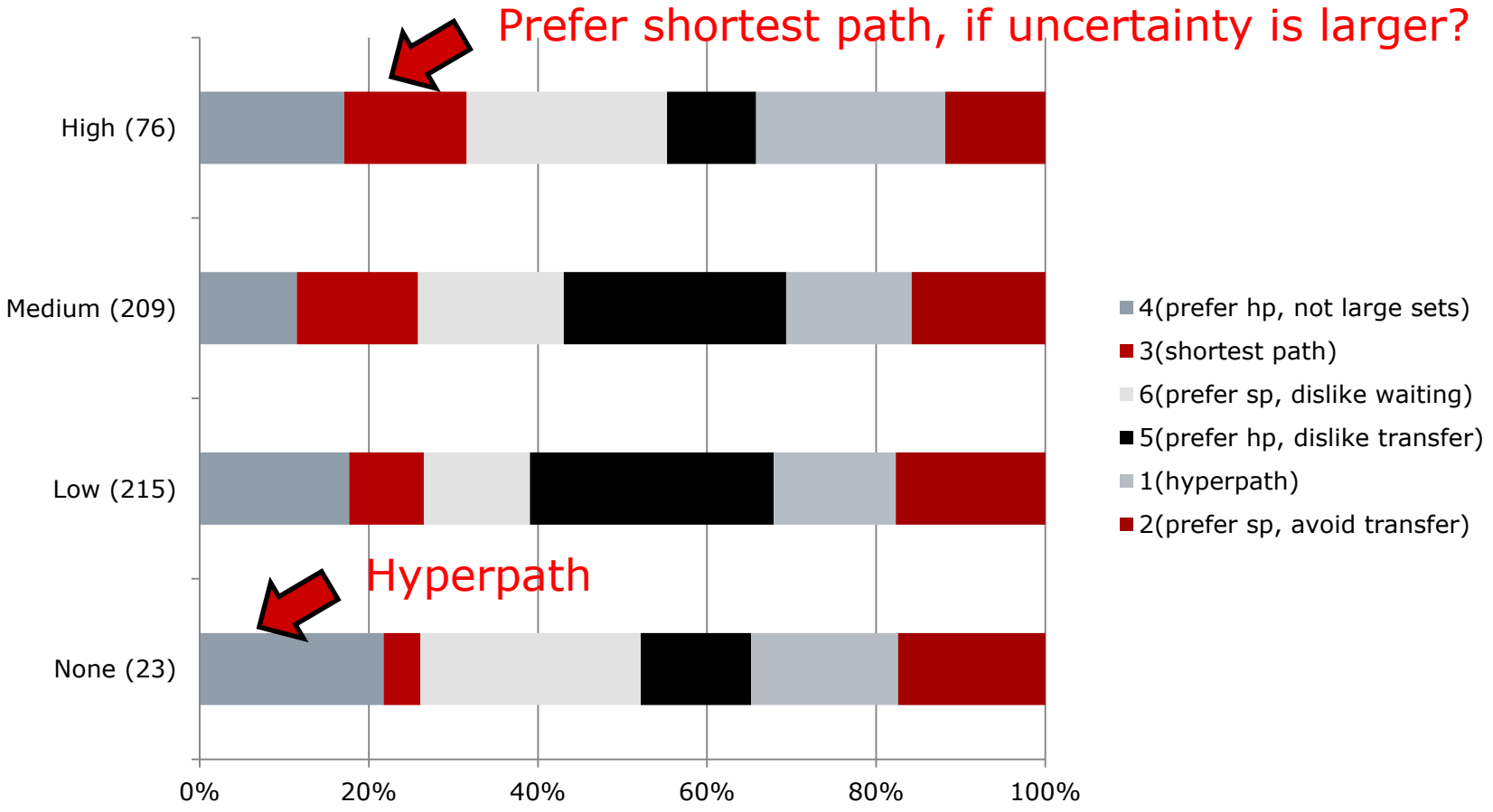
Behaviour clusters: age



Behaviour clusters: occupation



Behaviour clusters: travel time reliability



Conclusions

- Behavioural assumptions underpinning transport models might be excessively simplified
- International web-based survey to validate the S&F hyperpath models
- Travellers do consider strategies
  - Made smaller by information?
- Large number of behaviour clusters identified by SP experiments may point to a decision making process different from S&F's
  - But this is the conclusion of this SP experiments
- Can this variability be captured by changing the cost function or is a different heuristic needed?

Methodology: Stability-based validation

- The most problematic aspect is the choice of the number of clusters
- TwoStep clustering has some shortcomings
  - Sensitive to order of records
  - Heuristic
  - Goodness-of-fit evaluated through a geometrical validity measure, the Silhouette coefficient
    - Geometrical measures consider compactness, isolation, within and between-cluster dispersion, etc
    - A shape of clusters is assumed and results on the number of clusters are not reliable if actual clusters have a different shape
- A stability-based approach has been applied together with the SPSS standard procedure to overcome such limitations
  - *The true number of clusters is sought as the value for which the partitions obtained through data perturbation are highly similar to one another*
    - I.e. two objects are (not) in the same clusters regardless small perturbations of input data
  - NB: also stability might not be a good indicator of accuracy

Methodology: Validation approach

1. Specify the model setting the number of clusters n
2. Randomly split the sample in two disjoint subsamples S1 and S2 almost of the same size
3. Build a n cluster model of S1 using the SPSS TwoStep procedure, say C_1^{TwS}
4. Using the cluster distribution in C_1^{TwS} as dependent variable, build a Classification Tree and use it to predict a clustering of S1, say C_1^{CTr}, and S2, say C_2^{CTr}
   4.1. Evaluate the Misclassification Risk (MR) associated to the tree
5. Build a n cluster model of S2 using the SPSS TwoStep procedure, say C_2^{TwS}
6. Compare C_2^{CTr} with C_2^{TwS} using the Adjusted Rand Index (ARI)
7. Repeat steps from 2to 6 for 5 times in total

Capacity of the CT to understand the cluster rules

Stability of the cluster model

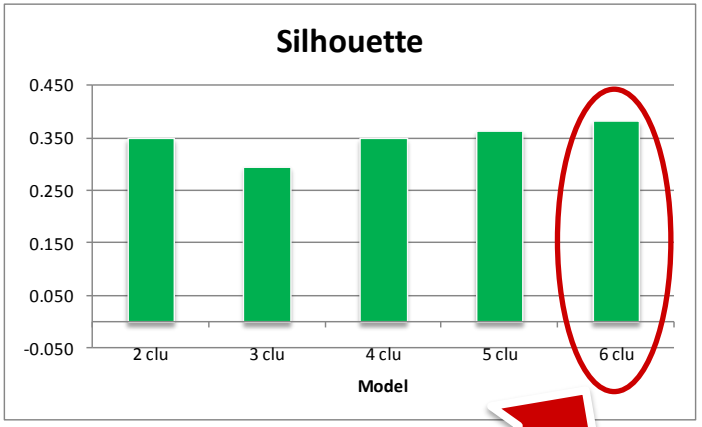
Methodology: Adjusted Rand Index (ARI)

- The similarity of the two clusterings of S_2 is measured by ARI
- The original Rand index is a measure of the similarity of two partitions of a set of objects
  - Let *P* and *Q* two partitions of a set *O* of *r* objects
  - a the number of objects which are in the same set both in *P* and in *Q*, *b* the number of objects which are in different sets both in *P* and in *Q*
  - The Rand index is the ratio (a+b)/bin.coef (*r*,2)
- ARI has been proposed to correct the fact that the expected value of the Rand index of two random partitions is not constant
  - ARI ranges between 0 and 1.
- Rand index and ARI are frequently used as measure of external validity of a clustering when correct clusters are known a priori
- In our procedure the clusters of S_2 generated by the rules underpinning C_1^{TwS} are assumed correct and compared with those generated by applying the SPSS TwoStep (i.e. C_2^{TwS})

Behaviour clusters: 6 clusters

3rd best model

Number of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Change[a]	Ratio of BIC Changes[b]	Ratio of Distance Measures[c]
1	4912.597			
2	4303.168	-609.429	1.000	1.464
3	3910.842	-392.326	.644	1.230
4	3606.064	-304.778	.500	1.436
5	3416.675	-189.389	.311	1.003
6	3227.952	-188.723	.310	1.420
7	3117.327	-110.625	.182	1.164
8	3032.913	-84.414	.139	1.115
9	2964.919	-67.994	.112	1.064
10	2905.504	-59.415	.097	1.124
11	2860.915	-44.589	.073	1.069
12	2824.049	-36.866	.060	1.032
13	2790.660	-33.390	.055	1.106
14	2767.707	-22.953	.038	1.016
15	2746.285	-21.422	.035	1.066



Best model