WEB APPENDIX FOR

IMPACT OF MENTAL REPRESENTATION ON CONSUMER BEHAVIORS: IMPLICATIONS FOR MENTAL BUDGETING AND PREDICTION ALGORITHM PREFERENCES

WEB APPENDIX FOR:

CONSUMERS' MENTAL REPRESENTATION OF EXPENDITURES: IMPLICATIONS FOR SPENDING AND SAVINGS DECISIONS

Table of Contents

APPENDIX A: List of Items

List of the 64 items used in categorization for all studies:

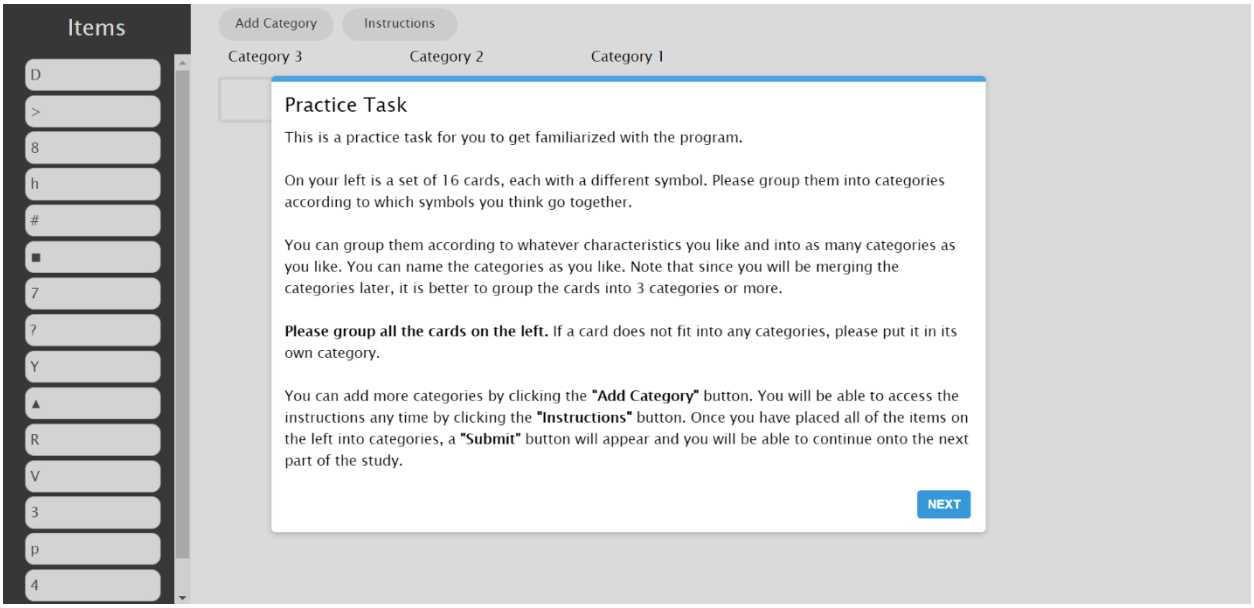
Rent	**Electricity**	**Car Maintenance**	**Internet**	**Mortgage**	**Insurance**	**Student Loans**	**Credit Card Debts**
Retirement savings	Phone	Storage	Emergency	Water	Gasoline	Shampoo	Milk
Bread	Chips	Soap	Medicine	Meats	Pet Food	Toilet Paper	Juice
Detergents	Chocolate	Coffee	Wine	Microwave	Dining Out	Uber	Train
Toll payments	Airplane tickets	Hotels booking	Movies	Pots and pans	Sport games	Going to bars	Birthday gifts to friends
Skates	Headphones	Vacuum Cleaner	Battery	Flowers	Video Games	Netflix	Toys
Laptop	Pizza	Gym Membership	Burger	Suitcase	Books	Ski Trips	Liquor
Souvenirs	Shoes	Jeans	Shirts	Watch	Sunglasses	Holiday Purchases	Weekend Getaway

Pilot study

The list of items was piloted on eight participants in the lab. The participants saw 64 index cards, each with a different item, and were asked to put together the items that go together. The pilot served 2 purposes: 1) to spot consistent outliers that were not categorized with any other items and 2) to probe from participants whether there were important expenditures that were missing from this set. From the pilot group, there was no consistent outliers that did not go together with other groups. Further, no other expenditure items were mentioned more than once by participants. Therefore, we proceeded to the main studies with this set of items.

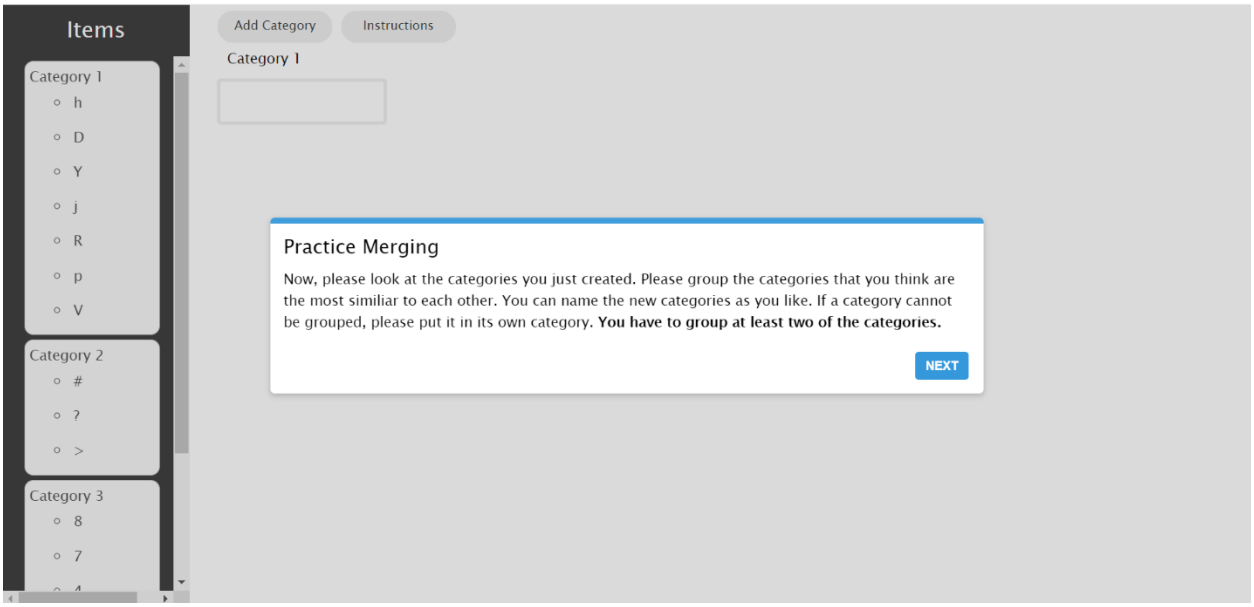
APPENDIX B: Successive Pile-sort Online Interface

Stage 1: Instructions for Practice Task – Initial Level Grouping



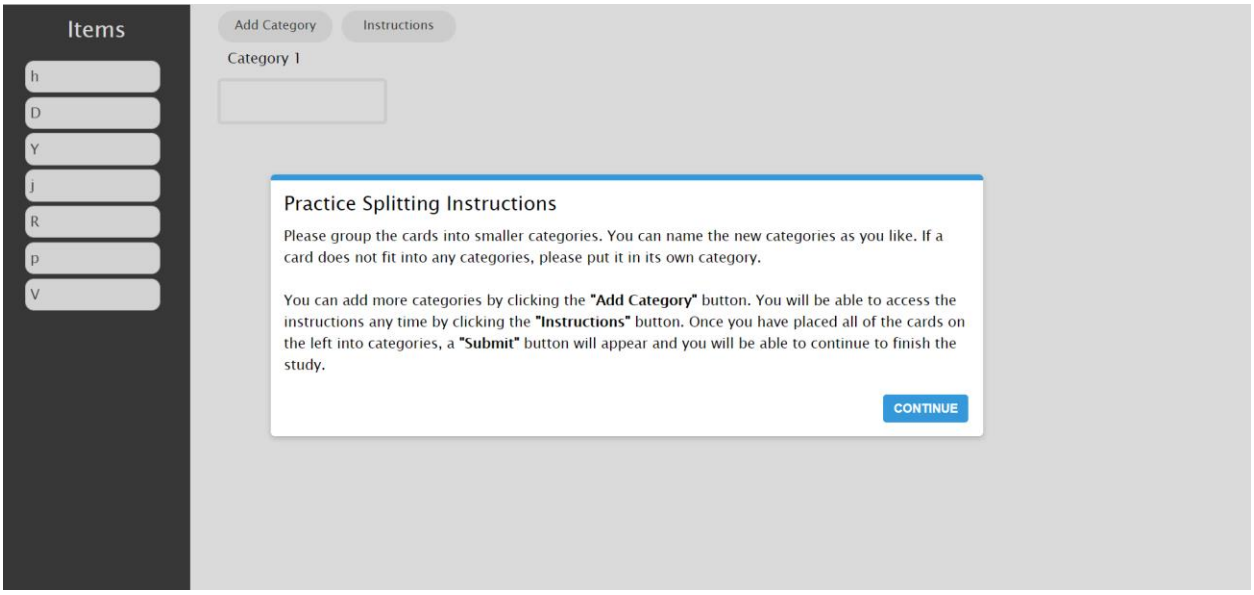
Item cards appear on the left, and the categories appear in the center. Participants can add, rename and delete categories as they wish. The categories constructed at this level are the initial level groupings.

Stage 2: Merging



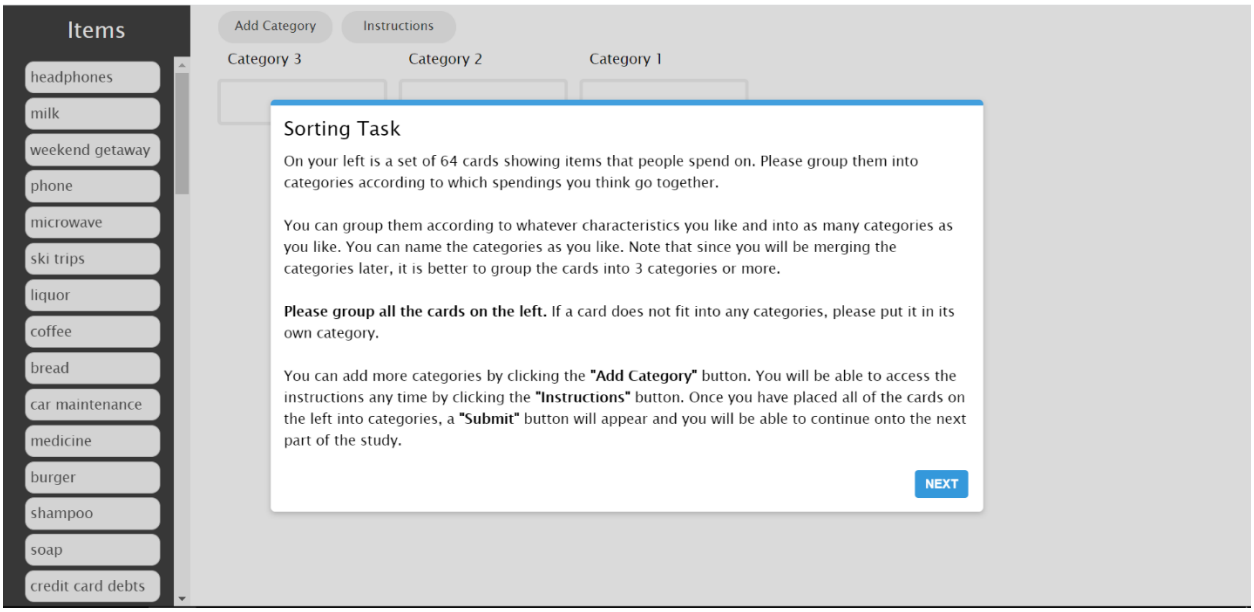
Categories that participants just construct will appear on the left, and they will drag the category cards to merge them into larger (higher-level) categories.

Stage 3: Splitting



All the items sorted into one category during the first sort (stage 1) appear on the left, and participants need to further split the items into smaller (lower-level) categories. Participants will go through all of their categories constructed at stage 1, one by one.

Main Task



After the practice task, participants will go through the same three stages for the main task with the 64 budgeting items (Appendix A).

APPENDIX C: Norming Data on the 64 Expenditures (Study 1 and 2)

	Wanting to Control [a]	Budget [a]	Hedonic [a]	Likelihood to Spend [b]		Amount Spent [b]	
Items	Mean	Mean	Factor Loading	Mean	SD	Mean	SD
rent	4.5	4.7	-4.6	5.6	2.0	743.1	690.4
electricity	4.5	5.0	-4.5	6.4	1.0	112.4	123.4
car maintenance	5.0	3.8	-4.0	5.6	1.5	113.7	158.4
internet	5.0	5.7	-3.5	6.5	0.9	74.5	90.5
mortgage	3.9	3.9	-3.2	5.4	1.9	577.2	585.5
insurance	5.0	4.6	-4.0	6.0	1.1	139.2	140.8
student loans	4.4	3.5	-1.9	4.5	2.2	787.4	1714.3
credit card debts	5.0	4.1	-0.8	5.6	1.6	252.5	262.2
retirement savings	5.0	3.9	-3.4	5.5	1.8	259.1	244.3
phone	4.7	5.3	-3.3	6.1	1.2	113.0	212.1
storage	3.8	3.4	-1.1	4.4	1.7	68.0	64.1
emergency	4.8	4.3	-4.0	5.7	1.3	134.4	324.5
water	4.3	4.5	-5.1	5.9	1.7	80.4	194.8
gasoline	5.1	4.4	-4.1	6.2	1.1	68.5	86.8
shampoo	4.1	4.1	-4.1	6.0	1.4	18.0	22.8
milk	4.2	4.0	-2.6	5.7	1.6	21.8	29.5
bread	4.1	4.0	-3.3	6.0	1.2	19.0	21.1
chips	4.3	3.6	1.6	5.3	1.5	19.7	29.1
soap	3.9	4.2	-4.8	6.0	1.3	15.5	15.2
medicine	4.6	4.4	-4.9	5.8	1.2	91.3	228.4
meats	4.8	3.7	-2.0	5.6	1.7	59.3	46.4
pet food	3.8	4.1	-3.4	5.3	2.1	40.5	38.3
toilet paper	4.0	4.3	-4.7	6.3	1.0	27.4	60.5
juice	3.7	3.5	-0.5	5.6	1.6	33.1	56.5
detergents	4.1	4.3	-4.0	5.7	1.5	19.6	23.7
chocolate	4.3	3.5	1.8	5.5	1.8	25.3	47.2
coffee	4.3	4.1	-0.1	5.8	1.7	25.5	28.4
wine	4.3	3.5	2.6	4.4	2.1	35.5	30.4
microwave	3.6	3.0	-2.3	4.9	1.8	62.7	132.5
dining out	5.3	5.0	2.9	5.1	1.8	103.6	127.6
uber	4.0	3.2	1.3	4.2	2.2	42.6	52.7
train	3.4	3.1	-0.4	3.6	1.9	117.5	294.6
toll payments	4.0	3.2	-1.8	5.1	1.8	22.4	28.9
airplane tickets	4.7	3.3	1.9	4.3	1.6	184.0	281.5
hotels booking	4.2	3.2	2.1	4.7	1.8	232.3	254.2
movies	4.1	3.7	2.4	5.5	1.8	39.2	65.4
pots and pans	3.9	3.0	-3.4	5.3	1.6	44.9	52.7
sport games	3.9	2.8	2.9	4.6	2.0	42.1	56.1

going to bars	4.1	3.1	3.8	3.6	1.9	77.6	115.5
birthday gifts to friends	4.7	4.4	0.9	5.3	1.9	35.8	26.5
skates	3.0	2.4	3.2	3.7	2.0	47.0	69.0
headphones	4.3	3.3	0.5	5.1	1.7	44.5	93.3
vacuum cleaner	4.1	3.1	-2.8	4.9	1.5	47.5	52.1
battery	4.0	3.1	-3.0	5.4	1.6	28.6	51.1
flowers	3.6	2.6	2.2	4.1	2.2	39.7	57.8
video games	4.0	3.5	2.8	5.3	1.4	40.9	34.5
netflix	3.9	4.8	2.4	5.9	1.4	26.2	64.6
toys	3.5	2.8	1.8	4.3	2.1	27.7	28.0
laptop	4.8	3.9	-1.6	5.4	1.8	264.5	316.9
pizza	4.3	4.0	0.9	5.8	1.2	30.2	32.0
gym membership	3.8	3.4	1.2	4.2	2.2	46.2	55.8
burger	4.4	3.3	0.4	5.0	2.0	23.2	24.4
suitcase	3.7	2.8	-0.4	4.3	1.9	44.6	82.6
books	4.2	3.5	-1.4	5.6	1.5	69.0	144.4
ski trips	3.5	2.6	4.3	3.5	2.2	45.5	47.0
liquor	4.7	3.5	2.8	4.6	2.1	30.7	29.9
souvenirs	3.8	3.1	3.2	4.7	1.4	41.7	83.1
shoes	4.4	3.9	-3.8	6.0	1.2	42.4	43.2
jeans	4.5	3.5	-2.5	5.5	1.6	54.4	76.0
shirts	4.6	4.0	-4.0	5.9	1.2	59.7	108.7
watch	3.8	2.9	1.2	4.4	2.1	51.3	59.6
sunglasses	4.0	2.6	-0.2	4.7	2.1	35.2	80.5
holiday purchases	4.5	4.2	2.3	5.5	1.5	78.0	69.7
weekend getaway	4.2	3.7	3.4	4.9	1.7	120.5	159.4

[a] To obtain this set of norming data, we recruited 297 Prolific workers and randomly assigned them to rate all 64 items on one of six dimensions (7-point likert scale). The dimensions are the following:

1) Frivolous: In your opinion, to what extent would you describe each of the items as **frivolous**? Frivolous products are something fun, enjoyable, experiential, and perhaps even "guilty" to be spending on. (1- Definitely Not; 7 - Definitely Yes)
2) Practical: In your opinion, to what extent would you describe each of the items as **practical**? Practical products are functional and useful, and are those that one ordinarily

buys to carry out a necessary function or task in one's life. (1- Definitely Not; 7 - Definitely Yes)

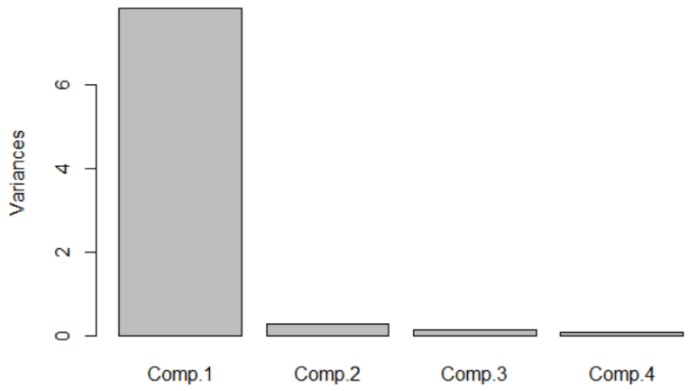
3) Luxury: In your opinion, to what extent would you describe each of the items as **a luxury**? Luxury products are something that are nice to own that you don't necessarily need to buy. (1- Definitely Not; 7 - Definitely Yes)
4) Necessity: In your opinion, to what extent would you describe each of the items as **a necessity**? Necessities are something that you need to buy. (1- Definitely Not; 7 - Definitely Yes)
5) Budget: In your opinion, to what extent do **you have a budget for each of the following items**? Having a budget means that you have a set spending amount for this item on a regular basis and rarely deviate away from it. (1- Definitely Not; 7 - Definitely Yes)
6) Control: In your opinion, to what extent do you **want to control spending for each of the following items**? This means that your current spending on this item fluctuates from time to time and you want to have a more rigid budget around it. (1- Definitely Not; 7 - Definitely Yes)

Examination of data reveals that the first four dimensions are highly correlated (table A1), and a screeplot (figure A1) suggests that they load on to the same component. Therefore, the first four dimensions are collapsed onto one factor that indicate how hedonic the spending is. The remaining two dimensions were kept as they were.

Correlation Between Norming Dimensions

	Frivolous	Practical	Luxury	Necessity	Budget	Control
Frivolous	1	-0.93	0.94	-0.911	-0.417	-0.267
Practical	-0.93	1	-0.884	0.94	0.458	0.288
Luxury	0.94	-0.884	1	-0.89	-0.409	-0.241
Necessity	-0.911	0.94	-0.89	1	0.583	0.392
Budget	-0.417	0.458	-0.409	0.583	1	0.648
Control	-0.267	0.288	-0.241	0.392	0.648	1

Screeplot of Principal Components on the First Four Dimensions



[b] To obtain this set of norming data, we recruited 59 Amazon Mechanical Turk workers and assigned them to either: 1) rate the likelihood to spend on all 64 items, or 2) indicate the amount that they typically spend on all 64 items. The likelihood question is phrased as the following: "Assume that each of the items was offered to you at a fair and reasonable price, would you ever consider spending on it?" The amount-to-spend question is phased as: "On average, if you were to spend on these items, how much would you spend on them every month?

APPENDIX D: Norming Study and Stimuli Selection for Study 2c

Norming Study

Stimuli

Item 1	Item 2	Distance
mortgage	rent	1
credit card debts	student loans	1
soap	shampoo	1
toilet paper	shampoo	1
detergents	shampoo	1
juice	milk	1
coffee	milk	1
meats	bread	1
chocolate	chips	1
pizza	chips	1
toilet paper	soap	1
detergents	soap	1
coffee	juice	1
liquor	wine	1
going to bars	dining out	1
weekend getaway	hotel booking	1
burger	pizza	1
weekend getaway	ski trips	1
jeans	shoes	1
shirts	shoes	1
shirts	jeans	1
mortgage	credit card debts	2
mortgage	toilet paper	3
mortgage	liquor	4
credit card debts	rent	2
credit card debts	toilet paper	3
credit card debts	liquor	4
soap	bread	2
soap	pizza	3
soap	ski trips	4
toilet paper	detergents	2
toilet paper	chocolate	3
toilet paper	ski trips	4
detergents	chocolate	3
detergents	ski trips	4
juice	burger	2
detergents	juice	3
juice	hotel booking	4
coffee	chocolate	2

detergents	coffee	3
coffee	ski trips	4
coffee	meats	2
meats	shirts	3
meats	ski trips	4
credit card debts	chocolate	4
juice	pizza	2
toilet paper	pizza	3
credit card debts	pizza	4
chocolate	liquor	2
liquor	going to bars	3
liquor	student loans	4
going to bars	weekend getaway	2
going to bars	wine	3
soap	going to bars	4
liquor	weekend getaway	3
weekend getaway	bread	4
soap	burger	3
credit card debts	burger	4
jeans	shampoo	2
toilet paper	jeans	3
jeans	hotel booking	4
shirts	shampoo	2
toilet paper	shirts	3
shirts	ski trips	4
toilet paper	rent	3
going to bars	rent	4
rent	student loans	2
toilet paper	student loans	3
shampoo	bread	2
pizza	shampoo	3
shampoo	ski trips	4
milk	chips	2
milk	shoes	3
milk	ski trips	4
coffee	bread	2
bread	shoes	3
bread	ski trips	4
detergents	chips	3
chips	hotel booking	4
juice	wine	2
mortgage	wine	4
weekend getaway	dining out	2
wine	dining out	3
soap	dining out	4
going to bars	hotel booking	2

liquor	hotel booking	3
bread	hotel booking	4
going to bars	ski trips	2
liquor	ski trips	3
shampoo	shoes	2
toilet paper	shoes	3
hotel booking	shoes	4

Participants and Procedures

We pre-registered to recruit 400 participants on Prolific. Each participant will rate either both the complement questions or both the substitute questions (See below) for 30 pairs of stimuli that are randomly selected out of 92 pairs of stimuli as shown above.

The 92 pairs include 21 pairs of items that are grouped together mostly at distance 1 and 2 and 25 pairs of items grouped together at distance 3 and 4.

Substitutes

1) Question 1: Assume that each pair of the items was offered to you at a fair and reasonable price. To what extent does one item satisfy your need for the other? (1 - One does not satisfy my need for the other at all; 7 - One completely satisfies my need for the other)
2) Question 2: If you wanted to spend on/buy one of them, but could only find the other, how likely would you be to spend on/buy the other one instead, assuming it is offered at a fair price? (1 - Definitely will not buy the other; 7 - Definitely will buy the other)

Complements

1) Question 1: Assume that each pair of the items was offered to you at a fair and reasonable price. To what extent do you think that one item is needed when you spend on/use the other? (1 - One is not needed at all when I use or consumer the other; 7 - I cannot use or consumer one without the other)

2) Question 2: To what extent do you think that one item provides something that the other lacks? (1 – Not at all; 7 – A lot)

Stimuli Selection

Of the normed stimuli, we selected three pairs that are very high (and low) on rated substitutability (and complementarity, resulting in a total of 12 product pairs. We used all the items in the 12 pairs as the focal item for study 2c, resulting in a total of 24 focal items (i.e., spending scenarios). This selection aimed to maximize the variance in the substitutability and complementarity ratings among the products that are often grouped together at the lowest level, so that we can best capture the effect of these cross-product relationships. The full stimuli are shown below.

Focal	Target	Distance (Mean)	Substitutability (Mean)	Complementarity (Mean)
Weekend getaway	Hotel booking	1.53	4.88	4.70
Weekend getaway	Going to bars	2.16	3.81	3.86
Weekend getaway	Liquor	3.21	3.60	4.00
Weekend getaway	Bread	3.70	2.13	3.26
Shirts	Jeans	1.31	4.12	4.98
Shirts	Shampoo	2.87	2.19	3.10
Shirts	Toilet Paper	2.94	1.82	3.30
Shirts	Ski Trips	3.53	2.10	4.60
Toilet paper	Shampoo	1.48	2.70	3.35
Toilet paper	Detergents	1.50	2.78	3.38
Toilet paper	Chocolate	3.00	1.87	3.20
Toilet paper	Ski trips	3.70	1.90	3.28
Detergents	Soap	1.41	5.03	4.10
Detergents	Toilet Paper	1.50	2.78	3.38
Detergents	Chocolate	3.01	2.06	2.77
Detergents	Ski Trips	3.68	1.87	3.03
Liquor	Wine	1.42	5.10	3.67
Liquor	Chocolate	2.23	2.91	3.37
Liquor	Going To Bars	2.87	4.46	4.65
Liquor	Student Loans	3.60	1.63	3.09
Burger	Pizza	1.37	5.09	3.18
Burger	Juice	2.04	3.12	3.82
Burger	Soap	3.02	2.08	3.12

	Credit Card			
Burger	Debts	3.53	2.27	3.07
Watch	Sunglasses	1.88	2.62	3.17
Watch	Jeans	2.37	2.49	3.08
	Holiday			
Watch	Purchases	3.08	2.63	2.58
Watch	Mortgage	3.62	1.71	1.29
Burger	Pizza	1.37	5.09	3.18
Burger	Juice	2.04	3.12	3.82
Burger	Soap	3.02	2.08	3.12
	Credit Card			
Burger	Debts	3.53	2.27	3.07
Coffee	Juice	1.51	3.71	3.20
Coffee	Chocolate	2.01	3.92	3.88
Coffee	Detergents	2.85	2.02	3.01
Coffee	Ski Trips	3.60	1.93	3.75
Shirts	Shoes	1.43	3.37	4.86
Shirts	Shampoo	2.87	2.19	3.10
Shirts	Toilet Paper	2.94	1.82	3.30
Shirts	Ski Trips	3.53	2.10	4.60
Juice	Milk	1.40	3.85	2.93
Juice	Burger	2.04	3.12	3.82
Juice	Detergents	2.82	1.93	2.79
Juice	Hotel Booking	3.64	1.90	3.05
Toilet paper	Soap	1.44	3.18	4.02
Toilet paper	Detergents	1.50	2.78	3.38
Toilet paper	Chocolate	3.00	1.87	3.20
Toilet paper	Ski trips	3.70	1.90	3.28
	Weekend			
Hotel booking	getaway	1.53	4.88	4.70
Hotel booking	Going to bars	2.35	2.37	3.46
Hotel booking	Liquor	3.29	2.31	3.53
Hotel booking	Shoes	3.53	1.92	3.27
Milk	Juice	1.40	3.85	2.93
Milk	Chips	1.99	2.60	3.17
Milk	Shoes	3.19	1.87	2.94
Milk	Ski Trips	3.71	1.56	3.25
Sunglasses	Watch	1.88	2.62	3.17
Sunglasses	Jeans	2.13	2.61	3.48
	Vacuum			
Sunglasses	Cleaner	2.94	1.78	1.19
Sunglasses	Mortgage	3.63	1.66	1.33
Soap	Toilet Paper	1.44	3.18	4.02
Soap	Bread	2.62	2.23	3.45
Soap	Burger	3.02	2.08	3.12
Soap	Ski Trips	3.71	1.69	3.67

Pizza	Burger	1.37	5.09	3.18
Pizza	Juice	2.05	3.16	3.75
Pizza	Shampoo	3.01	1.94	2.66
Pizza	Credit Card Debts	3.54	2.40	3.06
Jeans	Shirts	1.31	4.12	4.98
Jeans	Shampoo	2.94	2.03	3.18
Jeans	Toilet Paper	2.99	2.00	3.05
Jeans	Hotel Booking	3.50	1.62	3.10
Wine	Liquor	1.42	5.10	3.67
Wine	Juice	2.10	3.42	3.33
Wine	Going To Bars	2.98	4.83	4.25
Wine	Mortgage	3.58	2.13	3.17
Shampoo	Toilet Paper	1.48	2.70	3.35
Shampoo	Bread	2.65	1.90	3.03
Shampoo	Pizza	3.01	1.94	2.66
Shampoo	Ski Trips	3.70	1.80	3.42
Pizza	Burger	1.37	5.09	3.18
Pizza	Juice	2.05	3.16	3.75
Pizza	Shampoo	3.01	1.94	2.66
Pizza	Credit Card Debts	3.54	2.40	3.06
Juice	Coffee	1.51	3.71	3.20
Juice	Burger	2.04	3.12	3.82
Juice	Detergents	2.82	1.93	2.79
Juice	Hotel Booking	3.64	1.90	3.05
Soap	Detergents	1.41	5.03	4.10
Soap	Bread	2.62	2.23	3.45
Soap	Burger	3.02	2.08	3.12
Soap	Ski Trips	3.71	1.69	3.67
Shoes	Shirts	1.43	3.37	4.86
Shoes	Shampoo	2.86	1.82	3.08
Shoes	Milk	3.19	1.87	2.94
Shoes	Hotel Booking	3.53	1.92	3.27

APPENDIX E: Robustness checks for models in study 2

Study 2a Individualized Stimuli

	Dependent variable: Likelihood to adjust spending when overspent		
	Individual Random Effect	Individual Random Effect + Focal Controls	Individual Random Effect + Focal Controls + Individual Differences
Distance	0.104***	0.088***	0.326**
	(0.028)	(0.028)	(0.159)
FocalDiningOut		-0.150	0.039
		(0.306)	(0.301)
FocalMicrowave		-0.179	-0.047
		(0.356)	(0.348)
FocalShirts		-0.254	-0.081
		(0.300)	(0.296)
FocalSpending		-0.002	-0.004*
		(0.002)	(0.002)
Hedonic		-0.085***	-0.077**
		(0.033)	(0.032)
Budget		0.410***	0.396***
		(0.102)	(0.102)
Control		-0.183	-0.181
		(0.117)	(0.116)
TS_Score			0.016
			(0.028)
Short-termPlan			-0.158**
			(0.077)
Long-termPlan			-0.001
			(0.068)
Income			0.059**
			(0.028)
Distance:FocalSpending			0.001
			(0.001)

Distance:TS_Score			-0.009
			(0.008)
Distance:Income			-0.010
			(0.008)
Constant	2.957***	2.533***	2.120**
	(0.098)	(0.565)	(0.844)
Observations	644	644	644
Log Likelihood	-899.070	-892.040	-907.276
Akaike Inf. Crit.	1,806.141	1,806.079	1,850.552
Bayesian Inf. Crit.	1,824.012	1,855.224	1,930.971
Note:			*p<0.1; **p<0.05; ***p<0.01

	Dependent variable: Likelihood to adjust spending when underspent		
	(1)	(2)	(3)
Distance	0.003	-0.003	-0.078
	(0.021)	(0.021)	(0.120)
FocalDiningOut		-0.107	-0.107
		(0.209)	(0.214)
FocalMicrowave		0.139	0.167
		(0.254)	(0.259)
FocalShirts		0.025	0.028
		(0.208)	(0.214)
FocalSpending		0.001	0.002
		(0.001)	(0.001)
Hedonic		0.015	0.016
		(0.023)	(0.023)
Budget		0.219***	0.226***
		(0.083)	(0.083)
Control		0.189*	0.197**
		(0.097)	(0.098)
TS_Score			-0.012
			(0.021)
Short-termPlan			-0.040
			(0.057)

Long-termPlan			0.022
			(0.050)
Income			0.001
			(0.021)
Distance:FocalSpending			-0.001
			(0.0004)
Distance:TS_Score			0.007
			(0.006)
Distance:Income			0.001
			(0.006)
Constant	4.065***	2.525***	2.622***
	(0.070)	(0.409)	(0.637)
Observations	644	644	644
Log Likelihood	-697.547	-700.282	-724.818
Akaike Inf. Crit.	1,403.094	1,422.563	1,485.636
Bayesian Inf. Crit.	1,420.965	1,471.708	1,566.055
Note:			*p<0.1; **p<0.05; ***p<0.01

[a] Hedonic, Budget and Control are additional measures that we collected for each of the 64 items (See Appendix C).

[b] denotes a number of individual-difference measures that we collected and included in the robustness checks.

FocalSpending is the value that participants entered to indicate how much they usually budget for the focal item.

TS_Score is participants' tightwad-spendthrift scale value. Higher numbers indicate tendency to spend more (spendthrift).

Shorttermplan is participants' propensity-to-plan their financial resources in short term (a few days). Higher value indicate tendency to engage in more short-term financial planning.

Respectively, Longtermplan is participants' propensity-to-plan their financial resources in long term (1-2 months)

Income indicates participants' annual income.

Study 2b Aggregate stimuli

	Dependent variable: Likelihood to adjust when Overspending					
	(1)	(2)	(3)	(4)	(5)	(6)
Distance[a]	0.121***	0.239***				
	(0.035)	(0.026)				
AggDist[a]			0.001***			
			(0.0001)			
Dist_MDS				0.001***		
				(0.0001)		
Dist_HClust					0.001***	
					(0.0001)	
AggDist(Rank)						0.001***
						(0.0001)
OrderNormfirst	-0.041	0.004	0.039	0.039	0.039	0.039
	(0.090)	(0.075)	(0.044)	(0.044)	(0.044)	(0.044)
Scenario2	0.610***	-0.274***	-0.341***	-0.215***	-0.332***	-0.317***
	(0.070)	(0.086)	(0.095)	(0.092)	(0.095)	(0.095)
Scenario3	0.103	0.228***	0.134*	0.131*	0.173**	0.094
	(0.071)	(0.068)	(0.072)	(0.072)	(0.073)	(0.073)
Scenario4	0.112	-0.173**	-0.340***	-0.259***	-0.330***	-0.336***
	(0.070)	(0.074)	(0.081)	(0.079)	(0.081)	(0.082)
Scenario5	0.076	-0.254***	-0.382***	-0.286***	-0.328***	-0.359***
	(0.071)	(0.071)	(0.078)	(0.076)	(0.077)	(0.078)
Distance:OrderNormFirst	0.048					
	(0.051)					
Hedonic[b]		-0.134***	-0.143***	-0.130***	-0.140***	-0.135***
		(0.012)	(0.014)	(0.014)	(0.014)	(0.014)

Budget[b]		0.667***	0.617***	0.573***	0.700***	0.607***
		(0.061)	(0.070)	(0.070)	(0.070)	(0.070)
Control[b]		-0.297***	-0.242***	-0.165***	-0.290***	-0.222***
		(0.064)	(0.072)	(0.071)	(0.073)	(0.072)
Constant	2.539***	1.599***	1.151***	1.383***	0.997***	1.578***
	(0.077)	(0.201)	(0.217)	(0.218)	(0.217)	(0.224)
Observations	2,790	2,790	2,790	2.790	2.790	2.790
R^2			0.150	0.149	0.150	0.144
Adjusted R^2			0.147	0.146	0.147	0.142
Log Likelihood	-4,409.666	-4,208.602				
Akaike Inf. Crit.	8,839.331	8,441.204				
Bayesian Inf. Crit.	8,898.669	8,512.409				
Residual Std. Error			1.152 (df = 2780)	1.152 (df = 2780)	1.152 (df = 2780)	1.156 (df = 2780)
F Statistic			54.324*** (df = 9, 2780)	54.182*** (df = 9, 2780)	54.542*** (df = 9, 2780)	52.903*** (df = 9, 2780)

	Dependent variable: Likelihood to adjust when Underspending					
	(1)	(2)	(3)	(4)	(5)	(6)
Distance[a]	-0.075***	-0.101***				
	(0.028)	(0.023)				
AggDist[a]			-0.0003***			
			(0.0001)			
Dist_MDS				-0.0003***		
				(0.0001)		
Dist_HClust					-0.0003***	
					(0.0001)	
AggDist(Rank)						-0.0005***
						(0.0001)
OrderNormfirst	-0.010	-0.021	-0.019	-0.019	-0.019	-0.019

	(0.084)	(0.073)	(0.041)	(0.041)	(0.041)	(0.041)
Scenario2	0.071	-0.002	0.066	0.021	0.055	0.068
	(0.058)	(0.077)	(0.088)	(0.085)	(0.088)	(0.088)
Scenario3	0.371***	0.349***	0.474***	0.473***	0.460***	0.491***
	(0.059)	(0.061)	(0.067)	(0.067)	(0.068)	(0.067)
Scenario4	0.167***	0.208***	0.267***	0.242***	0.257***	0.276***
	(0.059)	(0.066)	(0.076)	(0.074)	(0.075)	(0.076)
Scenario5	-0.023	-0.040	0.029	-0.005	0.002	0.028
	(0.059)	(0.063)	(0.072)	(0.070)	(0.070)	(0.072)
Distance:OrderNormFirst	-0.011					
	(0.042)					
Hedonic[b]		0.004	0.008	0.006	0.005	0.007
		(0.011)	(0.013)	(0.012)	(0.013)	(0.013)
Budget[b]		0.219***	0.240***	0.264***	0.205***	0.247***
		(0.053)	(0.065)	(0.065)	(0.065)	(0.065)
Control[b]		0.043	0.020	-0.011	0.037	0.015
		(0.057)	(0.067)	(0.066)	(0.068)	(0.067)
Constant	4.008***	3.097***	3.238***	3.121***	3.298***	3.033***
	(0.069)	(0.178)	(0.201)	(0.202)	(0.201)	(0.206)
Observations	2,790	2,790	2,790	2,790	2,790	2,790
R^2			0.045	0.047	0.044	0.045
Adjusted R^2			0.042	0.044	0.041	0.042
Log Likelihood	-3,925.790	-3,904.257				
Akaike Inf. Crit.	7,871.579	7,832.513				
Bayesian Inf. Crit.	7,930.917	7,903.719				
Residual Std. Error			1.072 (df = 2780)	1.071 (df = 2780)	1.072 (df = 2780)	1.072 (df = 2780)
F Statistic			14.566*** (df = 9; 2780)	15.283*** (df = 9; 2780)	14.345*** (df = 9; 2780)	14.558*** (df = 9; 2780)
Note:				*p<0.1; **p<0.05; ***p<0.01		

[a] Distance is the dummy coded taxonomic distance: (0 = close-distance; 1= mid-distance; 2 = far-distance). AggDist is the aggregate level distance, obtained from the categorization data, of each pair of focal-comparison products. Dist_MDS is the distance recovered after multidimensional scaling was applied. Dist_HClust is the distance recovered after the hierarchical clustering was applied. AggDist(Rank) is the ordinal aggregate distance obtained by taking the ranks of the aggregate level pair-wise distance. The aggregate level distance is highly correlated with the distance recovered from the MDS (Pearson's $r = 0.95$, $t(2014) = 143.41$, $p < .001$), and hence additional robustness check using the MDS distance were not included.

[b] Hedonic, Budget and Control are additional measures that we collected for each of the 64 items (See Appendix C).

Study 2c Substitutes and Complements

	Different Specifications of the Models					
	Overspending Controlling for Both	Underspending Controlling for Both	Overspending Controlling for Substitutability	Overspending Controlling for Complementarity	Underspending Controlling for Substitutability	Underspending Controlling for Complementarity
Dist.1	0.152***	-0.112***	0.167***	0.173***	-0.107***	-0.081***
	(0.036)	(0.030)	(0.036)	(0.022)	(0.030)	(0.019)
Substitutability	-0.022	-0.032	-0.052*		-0.044*	
	(0.030)	(0.025)	(0.029)		(0.024)	
Complementarity	-0.120***	-0.044*		-0.125***		-0.052**
	(0.028)	(0.025)		(0.027)		(0.024)
Hedonic	-0.089***	-0.022**	-0.089***	-0.092***	-0.022**	-0.028***
	(0.011)	(0.009)	(0.011)	(0.009)	(0.009)	(0.008)
Budget	0.681***	0.092*	0.715***	0.671***	0.108**	0.078
	(0.061)	(0.054)	(0.061)	(0.060)	(0.053)	(0.053)
Control	-0.448***	0.178***	-0.447***	-0.455***	0.179***	0.169***
	(0.052)	(0.045)	(0.052)	(0.052)	(0.045)	(0.044)
Constant	-1.649***	-0.537**	-2.142***	-1.688***	-0.722***	-0.587**
	(0.267)	(0.233)	(0.242)	(0.262)	(0.209)	(0.230)
Observations	4,368	4,656	4,368	4,368	4,656	4,656
R^2	0.249	0.039	0.245	0.249	0.039	0.039
Adjusted R^2	0.248	0.038	0.245	0.248	0.038	0.038
Residual Std. Error	1.038 (df = 4361)	0.913 (df = 4649)	1.040 (df = 4362)	1.038 (df = 4362)	0.913 (df = 4650)	0.913 (df = 4650)
F Statistic	240.570*** (df = 6; 4361)	31.697*** (df = 6; 4649)	283.823*** (df = 5; 4362)	288.603*** (df = 5; 4362)	37.381*** (df = 5; 4650)	37.690*** (df = 5; 4650)

Note: *p<0.1; **p<0.05; ***p<0.01

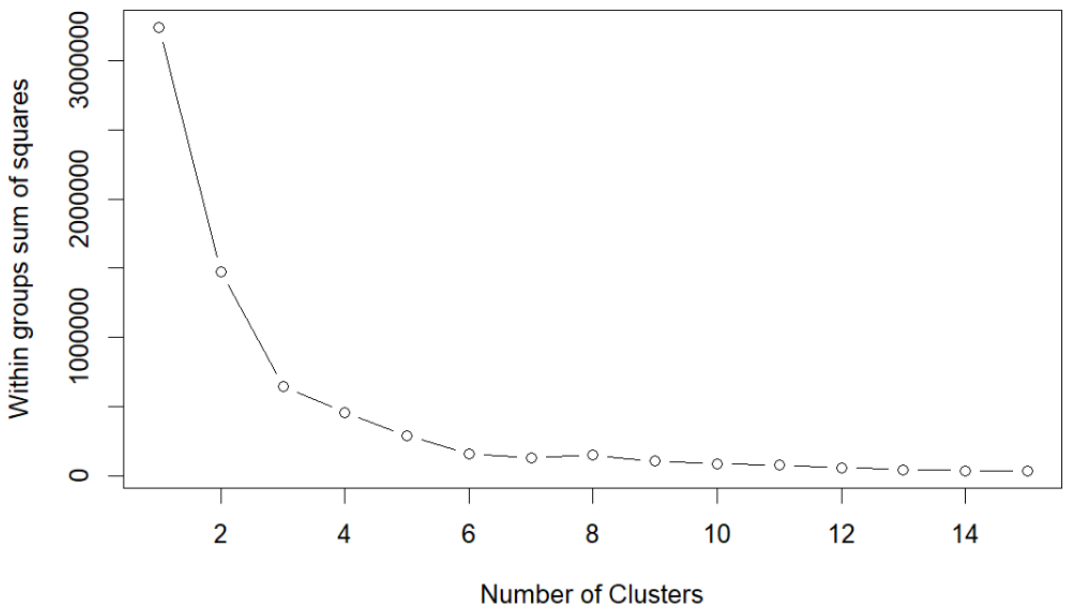
APPENDIX F: Stimuli Selection of Study 3

Study 3 aimed to test how participants adjust their spending in when items are on sale. Therefore, we used a range of products that are obtainable in a single shopping trip so that it is natural for participants to think about these products together with one another in the event of promotions. To generate the products, we first chose 45 items from a wholesale store's website. Then, we recruited 198 MTurk workers to perform successive pile sort on the 45 items. On aggregate, we classified the stimuli into four clusters as suggested by within group sum-of-squares (Figure A2). However, it turned out that the size of the 4 clusters were largely uneven (Figure A3), and therefore we added products that we expected to expand the smaller groups and cut products from larger groups.

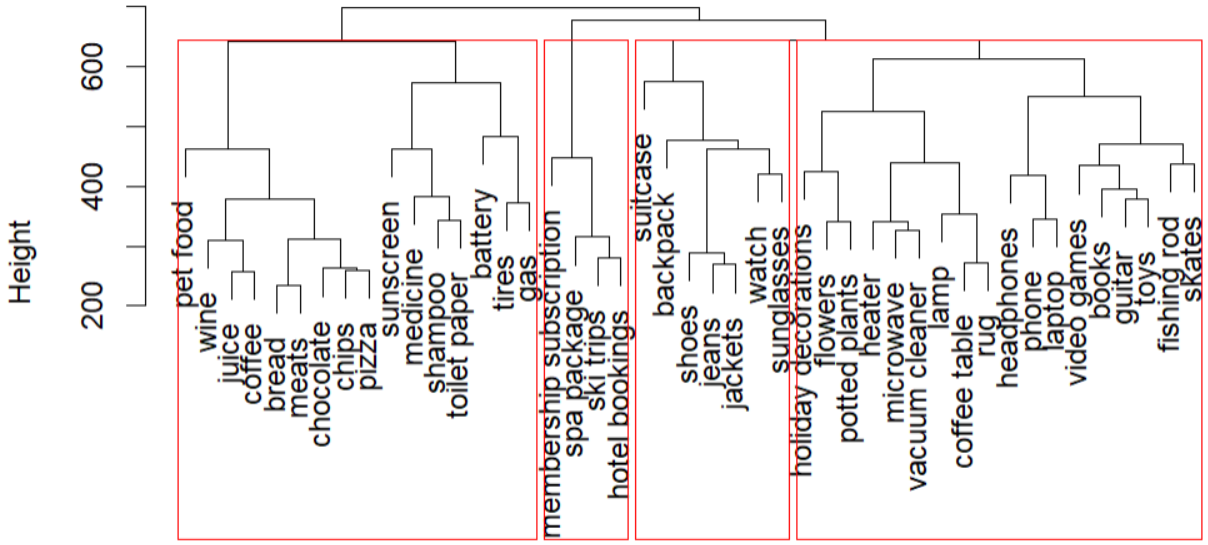
Initial Chosen Products (45 Items)

Phone	Meats	Fishing Rod	Suitcase	Potted Plants
Sunscreen	Pet Food	Skates	Books	Tires
Juice	Toilet Paper	Headphones	Shoes	Ski Trips
Backpack	Lamp	Vacuum Cleaner	Jeans	Hotel Bookings
Shampoo	Holiday Decorations	Battery	Jackets	Gas
Coffee Table	Chocolate	Flowers	Watch	Rug
Bread	Coffee	Video Games	Sunglasses	Heater
Chips	Wine	Guitar	Laptop	Membership Subscription
Medicine	Microwave	Toys	Pizza	Spa Package

Figure A2: Within Group Sum of Squares by the Number of Clusters



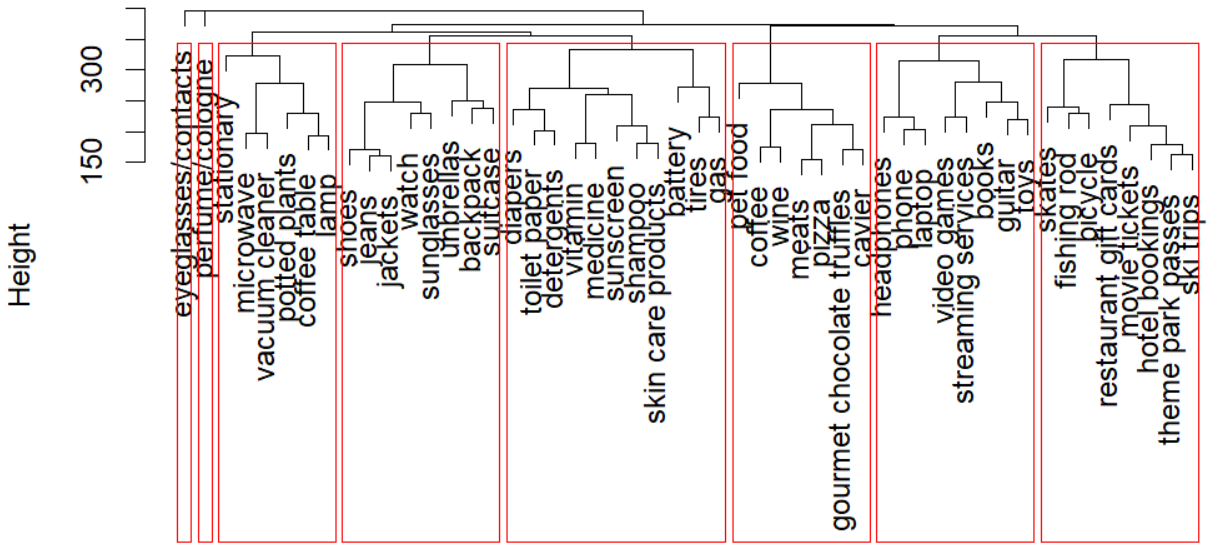
Dendrogram of Product Groupings



After changing the products, we ran a separate pilot on the new set of stimuli with 50 products with 99 MTurk workers. The mode of the initial number of categories was 6, and the set

now had more balanced groupings (Figure A4). Therefore, we proceed with the set with 50 products. The specific products in the stimuli were presented below.

Dendrogram of Product Groupings (Updated Stimuli)



Final Products Used as Stimuli for Study 3 (50 Items)

Phone	Pet Food	Headphones	Suitcase	Gas
Sunscreen	Toilet Paper	Vacuum Cleaner	Books	Perfume/Cologne
Theme Park Passes	Lamp	Battery	Shoes	Stationery
Backpack	Detergents	Skin Care Products	Jeans	Movie Tickets
Shampoo	Gourmet Chocolate Truffles	Video Games	Jackets	Restaurant Gift Cards
Coffee Table	Coffee	Guitar	Watch	Bicycle
Diapers	Wine	Toys	Sunglasses	Streaming Services
Vitamin	Microwave	Laptop	Tires	Umbrellas
Medicine	Fishing Rod	Pizza	Ski Trips	Caviar
Meats	Skates	Potted Plants	Hotel Bookings	Eyeglasses/Contacts

APPENDIX G: Norming Data on the 50 Promotion Items (Study 3 and 4)

Norming on Spending Related Dimensions for Each Product

	Typicality[a]		Spending likelihood[b]		Frivolous[b]		Necessity[b]		Control[b]		Budget[b]		Promotion likelihood[b]		Spending Amount[b]	
Items	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Phone	5.83	1.41	5.36	1.84	3.56	1.85	5.99	1.50	5.19	1.82	5.20	1.92	5.31	1.65	508.23	342.63
Sunscreen	5.14	1.72	4.34	2.11	2.20	1.35	4.79	2.00	4.13	1.81	3.83	2.07	4.07	1.88	9.28	5.48
Theme Park Passes	5.38	1.90	3.13	2.00	6.41	1.12	1.25	0.76	5.10	2.16	3.53	2.32	4.80	1.84	101.41	121.57
Backpack	4.76	1.77	4.13	2.14	3.10	1.69	3.81	1.99	4.51	1.94	3.74	2.19	4.44	1.64	40.81	25.80
Shampoo	5.96	1.41	5.69	1.69	2.06	1.42	6.09	1.36	4.65	1.90	4.79	1.91	4.94	1.74	8.86	5.81
Coffee Table	4.99	1.94	3.44	2.02	3.74	1.81	2.31	1.56	4.67	2.01	3.82	2.27	3.40	1.93	145.03	100.31
Diapers	4.80	1.91	2.16	2.02	1.94	1.55	4.96	2.29	4.00	2.03	2.64	2.05	4.00	1.97	21.54	13.32
Vitamin	5.30	1.70	4.89	1.85	2.93	1.85	4.88	1.85	4.48	1.84	4.14	2.08	4.34	1.94	16.23	13.96
Medicine	5.62	1.54	5.66	1.48	1.73	1.45	6.69	0.85	4.19	1.97	4.05	2.07	2.40	1.82	46.38	51.42
Meats	6.11	1.23	5.24	1.97	2.83	1.58	5.22	1.93	4.49	1.71	4.73	1.94	4.17	1.78	20.62	17.83
Pet Food	4.71	1.95	4.27	2.47	2.10	1.58	5.49	2.01	4.07	2.06	4.00	2.29	4.51	1.89	24.09	16.98
Toilet Paper	5.88	1.42	5.96	1.68	1.64	1.36	6.57	1.08	4.04	1.79	4.32	2.21	3.87	2.20	10.30	6.17
Lamp	5.24	1.72	4.21	1.90	2.90	1.64	4.46	1.95	4.57	1.74	3.79	2.12	3.81	1.85	37.29	24.06
Detergents	5.71	1.49	5.54	1.79	1.84	1.24	6.15	1.16	4.41	1.97	4.59	2.02	4.61	1.81	11.75	7.07
Gourmet Chocolate Truffles	5.03	2.05	3.30	2.05	6.24	1.41	1.46	1.16	4.80	2.11	3.39	2.29	3.69	1.80	24.02	27.19
Coffee	6.02	1.35	4.93	2.39	3.61	1.88	3.63	2.26	4.84	1.91	4.59	2.17	4.79	1.90	8.52	7.15
Wine	5.39	1.66	3.69	2.41	5.51	1.63	1.88	1.63	4.81	1.99	3.36	2.20	4.11	1.88	21.09	13.06
Microwave	5.23	1.73	4.16	2.05	2.77	1.60	4.46	1.89	4.86	1.81	4.03	2.23	4.06	1.93	111.00	95.41
Fishing Rod	4.69	2.03	2.56	2.03	4.80	1.71	2.07	1.62	4.49	2.24	3.20	2.27	3.71	1.76	66.90	45.98
Skates	4.60	2.00	2.54	1.95	5.19	1.76	1.49	0.92	4.59	2.05	2.92	2.09	3.60	1.86	59.84	37.20
Headphones	5.57	1.35	5.20	1.75	4.31	1.77	3.09	1.99	5.01	1.92	4.17	2.26	5.11	1.52	71.35	78.88
Vacuum Cleaner	5.41	1.66	4.07	2.12	2.66	1.68	4.62	1.97	4.80	1.85	4.00	2.08	4.36	1.93	115.80	75.01
Battery	5.02	1.66	4.94	1.86	2.07	1.37	5.63	1.41	4.35	1.89	3.73	2.12	3.73	1.88	21.41	29.61
Skin Care Products	5.69	1.54	4.60	1.91	3.70	1.79	4.41	1.79	4.88	1.81	4.61	1.98	4.87	1.73	31.87	60.74
Video Games	5.61	1.58	4.70	2.18	5.47	1.77	1.87	1.34	4.88	2.08	4.33	2.17	5.23	1.51	62.81	66.52
Guitar	4.89	1.98	2.66	1.97	5.34	1.26	1.82	1.51	4.61	2.18	3.26	2.26	3.81	1.84	255.74	208.73
Toys	5.26	1.77	3.44	2.14	4.63	1.81	2.29	1.58	4.62	1.99	3.50	2.09	5.19	1.80	21.52	13.93

Laptop	5.78	1.43	5.37	1.71	3.81	1.72	4.57	1.98	5.29	1.72	4.94	2.09	5.13	1.58	605.90	357.10
Pizza	5.78	1.56	5.36	1.79	4.01	1.81	2.50	1.77	4.64	1.82	4.38	2.02	5.43	1.64	15.75	7.15
Potted Plants	4.59	1.70	4.24	2.12	4.23	1.53	2.18	1.60	4.30	1.87	3.50	2.10	3.21	1.55	18.93	12.75
Suitcase	4.70	1.84	3.30	2.02	3.43	1.75	2.97	1.74	4.43	1.95	3.35	2.12	4.19	1.83	75.78	48.70
Books	5.29	1.53	5.06	1.82	2.97	1.70	4.34	1.87	4.54	1.74	4.21	2.02	4.50	1.62	15.77	8.60
Shoes	5.78	1.61	5.74	1.34	2.99	1.89	6.50	1.03	5.00	1.86	4.95	1.78	5.13	1.61	62.62	32.04
Jeans	5.82	1.57	5.17	1.70	3.23	1.58	4.69	1.96	4.84	1.90	4.71	1.95	5.00	1.58	40.97	20.28
Jackets	5.74	1.61	5.29	1.53	2.90	1.75	5.74	1.54	4.88	1.78	4.50	2.08	4.63	1.80	65.14	42.27
Watch	5.19	1.68	3.76	2.22	4.61	1.53	2.59	1.75	4.77	1.96	3.70	2.31	4.56	1.70	112.22	134.92
Sunglasses	5.08	1.65	4.44	2.05	4.26	1.46	3.56	1.97	4.81	1.83	4.14	2.07	4.63	1.70	43.84	46.88
Tires	4.78	1.95	4.43	2.10	1.94	1.44	5.88	1.60	4.59	1.93	4.20	2.16	4.27	2.01	292.71	315.50
Ski Trips	5.13	2.10	2.80	2.05	6.47	1.16	1.25	0.82	4.93	2.18	3.36	2.42	4.30	2.09	725.12	808.08
Hotel Bookings	5.36	1.70	3.86	2.09	5.23	1.76	1.85	1.40	4.99	2.03	3.80	2.33	4.93	1.94	152.61	95.69
Gas	5.30	1.88	5.81	1.80	1.80	1.41	6.40	1.19	4.38	1.96	4.47	2.21	2.63	1.93	20.97	16.64
Perfume (Cologne)	5.33	1.68	3.51	1.99	5.60	1.32	2.16	1.62	5.01	1.93	3.85	2.30	4.29	1.81	53.26	33.99
Stationery	4.43	1.95	3.67	2.01	4.04	1.79	2.69	1.64	4.55	1.99	3.53	2.24	3.86	1.64	14.43	19.62
Movie Tickets	5.61	1.58	3.83	2.04	5.59	1.54	1.65	1.40	4.84	2.06	3.52	2.25	3.90	2.08	16.43	8.67
Restaurant Gift Cards	5.08	1.77	3.50	2.11	5.44	1.69	1.56	1.01	4.90	2.09	3.62	2.12	3.66	2.06	36.96	21.48
Bicycle	4.98	1.77	3.96	2.13	3.67	1.59	3.31	1.74	4.52	1.88	3.50	2.19	4.23	1.76	213.93	186.12
Streaming Services	5.59	1.70	5.07	2.01	4.66	1.81	2.85	2.02	5.35	1.77	5.09	2.02	4.67	1.92	22.26	25.19
Umbrellas	4.59	1.78	3.46	1.86	2.67	1.61	3.75	1.90	4.20	2.14	3.24	1.98	3.47	1.72	15.52	7.80
Caviar	4.64	2.19	1.56	1.29	6.34	1.47	1.21	0.72	4.38	2.28	2.94	2.27	2.06	1.63	74.96	59.25
Eyeglasses (Contacts)	5.16	1.72	4.34	2.38	1.74	1.25	6.25	1.50	4.62	1.86	4.17	2.10	3.54	1.85	151.03	110.71

[a] To obtain item typicality ratings, we recruited 60 prolific workers to perform the successive pile sort on the 50 items while also rating the typicality of each item to the initial level category that they constructed.

[b] To obtain this set of norming data, we recruited 161 Prolific workers and randomly assigned them to answer three out of seven questions for all 50 items. The dimensions are the following:

1) Spending Likelihood: Assume that each of the items was offered to you at a fair and reasonable price, would you ever consider spending on it? Please indicate **how likely you are to spend on each of the following items.** (1- Extremely Unlikely; 7- Extremely Likely)
2) Spending Amount: For each item, **what do you think the retail price would be?**
3) Frivolous: In your opinion, to what extent would you describe each of the items as **frivolous**? Frivolous products are products that are fun, enjoyable, experiential, and are those where you might even feel "guilty" to spend money on them. (1- Definitely not frivolous; 7- Definitely frivolous)
4) Necessity: In your opinion, to what extent would you describe each of the items as **a necessity**? Necessities are something that you need to buy. (1- Definitely not; 7- Definitely yes)
5) Budget: In your opinion, to what extent do **you have a budget for each of the following items**? Having a budget means that you have a spending limit in mind for this item and that you try not to exceed this limit. (1- Definitely not; 7- Definitely yes)
6) Control: To what extent do you **want to control your spending** for each of the following items? This means that you want to keep your spending under control and have a firm budget for the item. (1- Definitely not; 7- Definitely yes)

7) Promotion Likelihood: In your opinion, how likely do you think each of these items would **be offered on a promotion**? Examples of promotions include being on sale for a price discount, coupons on the item, reduced priced when bundled with other products, and free gifts for purchasing the item. (1- Extremely Unlikely; 7- Extremely Likely)

Norming on the Demand Relationship Between Product Pairs

			Substitute Rating		Complement Rating	
Relationship	**Focal**	**Comparison**	**Mean**	**SD**	**Mean**	**SD**
Close	coffee	pizza	2.81	2.11	1.76	1.33
Far	coffee	stationery	2.23	2.05	1.70	1.30
Mid	coffee	detergents	2.16	1.96	1.42	1.10
Close	diapers	detergents	2.39	2.15	2.03	1.60
Far	diapers	wine	1.91	1.93	1.24	0.76
Mid	diapers	gas	2.00	2.00	1.34	0.94
Close	gourmet chocolate truffles	wine	3.41	2.27	2.90	1.79
Far	gourmet chocolate truffles	streaming services	2.20	1.90	2.01	1.55
Mid	gourmet chocolate truffles	pet food	2.09	2.06	1.23	0.68
Close	jeans	backpack	2.48	2.04	2.59	1.76
Far	jeans	restaurant gift cards	2.11	1.90	1.58	1.15
Mid	jeans	lamp	1.84	1.77	1.18	0.54
Close	lamp	pots and pans	2.23	2.04	1.82	1.37
Far	lamp	sunglasses	2.08	2.03	1.44	1.08
Mid	lamp	books	2.28	2.01	3.46	1.95
Close	meats	wine	2.76	2.03	2.80	1.75
Far	meats	toys	1.85	1.82	1.18	0.66
Mid	meats	gas	2.10	1.97	1.72	1.41
Close	movie tickets	streaming services	4.03	2.35	1.87	1.59
Far	movie tickets	diapers	1.83	1.83	1.24	0.85
Mid	movie tickets	toys	2.65	2.07	1.28	0.66
Close	perfume(cologne)	skin care products	3.39	2.25	2.90	1.78
Far	perfume(cologne)	lamp	1.86	1.85	1.30	0.85
Mid	perfume(cologne)	backpack	2.03	1.89	1.30	0.60
Close	pet food	pizza	2.17	2.02	1.32	0.87
Far	pet food	toys	2.18	1.98	1.90	1.51
Mid	pet food	diapers	1.93	1.92	1.24	0.73
Close	potted plants	pots and pans	2.26	2.09	1.62	1.47
Far	potted plants	sunglasses	2.02	1.94	1.52	1.14
Mid	potted plants	books	2.42	2.07	1.55	1.04
Close	restaurant gift cards	streaming services	2.63	2.09	1.49	1.07
Far	restaurant gift cards	pet food	1.96	1.92	1.31	0.93
Mid	restaurant gift cards	gourmet chocolate truffles	3.14	2.20	1.83	1.46
Close	shampoo	sunscreen	2.37	2.02	1.83	1.52
Far	shampoo	stationery	2.04	2.00	1.27	0.84

Mid	shampoo	coffee	2.06	1.95	1.28	0.91
Close	sunglasses	backpack	2.40	2.06	2.34	1.66
Far	sunglasses	restaurant gift cards	2.08	1.94	1.46	0.97
Mid	sunglasses	lamp	2.04	1.94	1.31	0.87
Close	toilet paper	detergents	2.54	2.10	1.97	1.59
Far	toilet paper	pizza	2.14	2.01	2.08	1.76
Mid	toilet paper	stationery	2.14	2.01	1.58	1.37

To obtain this set of norming data, we recruited 149 MTurk workers to rate the substitutive and complementary relationships for 42 product pairs (selected from minimizing the difference between the norming dimensions). There were two questions for the substitutive relationship and two questions for the complementary relationship (See question wordings below). Participants rated one substitute and one complement question, which were randomly chosen from the two.

Substitutes

3) Question 1: Assume that each pair of the items was offered to you at a fair and reasonable price. **To what extent does buying one item satisfy your need for the other?**
4) Question 2: **If you wanted to buy one of them at the store, but could only find the other, how likely would you be to buy the other one instead, assuming it is offered at a fair price?**

Complements

3) Question 1: Assume that each pair of the items was offered to you at a fair and reasonable price. **To what extent do you think that one item is needed when you use or consume the other?** (1- One is not needed at all when I use or consumer the other; 7 – I cannot use or consumer one without the other)

4) Question 2: Assume that each pair of the items was offered to you at a fair and reasonable price. **How often do you buy one with the other?** (1 – I never buy these items together; 7 – I always buy these items together)

The two measures of substitutes yielded a correlation of 0.92, while the measures of complements had only 0.52. Upon discussion, the authors decide to only use the first complement rating as the measure for complementary relationship, as the second question 1) does not mean that the two items have to be consumed together and 2) can also influence people's taxonomic distance. Therefore, the ratings in the table above are the average across two measures of substitutes (N = 149) and the average of the first question of complements (N = 73).

APPENDIX H: Robustness checks for models for study 3a

Predictors	Aggregate Distance	Aggregate Distance Controlling for Demand Relationships	Dendrogram Distance	Dendrogram Distance Controlling for Demand Relationships
Aggregate Distance	.005***	.005***		
	(.0008)	(.0014)		
Distance From Dendrogram			.007***	.007***
			(.0005)	(.0008)
Typicality	-0.49*	-0.56*	0.52	0.14
	(0.30)	(0.32)	(0.34)	(0.36)
Magnitude	0.23	0.20	-.030	-.039
	(0.30)	(0.32)	(0.34)	(0.36)
Distance*Typicality	.0027*	.0034*	-.002	-.0004
	(0.0016)	(0.0018)	(0.001)	(0.001)
Distance*Magnitude	-.0012	-.00111	.0001	.0002
	(0.0016)	(0.0018)	(0.001)	(0.001)
Intercept y = 2	-0.18	-0.11	-1.30***	-1.76***
	(0.15)	(0.45)	(0.17)	(0.38)
Intercept y = 3	-1.59***	-1.56***	-2.73***	-3.23***
	(0.15)	(0.45)	(0.18)	(0.38)
Substitutability[1]		0.0003		0.08
		(0.08)		(0.07)
Complementarity		-0.04		0.08
		(0.06)		(0.06)

Note: Number in parenthesis denote SE. *p<0.1; **p<0.05; ***p<0.01

[1] See Appendix G for how substitutability and complementarity are measured.

WEB APPENDIX FOR:

PREDICTION BY REPLICATION: PEOPLE PREFER PREDICTION ALGORITHMS THAT REPLICATE THE EVENT BEING PREDICTED

Table of Contents

APPENDIX A: Stimuli for Study 1b

Condition	Stimuli
Probability Match	This algorithm draws a marble from an opaque jar. Seven marbles of different colors are in the jar. Each color corresponds to a value between 1 and 7. The algorithm draws one marble and submits the corresponding value as its prediction. The colors and values are listed below: Blue: value = 1 Green: value = 2 Red: value = 3 Yellow: value = 4 Black: value = 5 Purple: value = 6 Orange: value = 7
Action Match	This algorithm rolls an 11-sided die with the values 1, 2, 3, 3, 4, 4, 4, 3, 5, 6, 7 on each side. The algorithm submits the die-roll outcome as its prediction.
Perfect Match	This algorithm rolls a 7-sided die with the values 1, 2, 3, 4, 5, 6, 7 on each side. The algorithm submits the die-roll outcome as its prediction.
No Match	This algorithm draws a marble from an opaque jar. Eleven marbles of different colors are in the jar. Each color corresponds to a value between 1 and 7. The algorithm draws one marble and submits the corresponding value as its prediction. The colors and values are listed below: Blue: value = 1 White: value = 2 Green: value = 3 Brown: value = 3 Red: value = 4 Yellow: value = 4 Black: value = 4 Purple: value = 5 Gray: value = 5 Orange: value = 6 Pink: value = 7
Constant	This algorithm always predicts the outcome of the die roll will be 4.

APPENDIX B: Pretest for Studies 2a

In this pretest, we recruited 199 MTurk workers (216 responded, 199 passed attention check). They read about the follwing algorithms (in randomized order):

Method E:

Method E is based entirely on team metrics (i.e. it doesn't account for individual player metrics). A given team's rating is generated by aggregating past team data like margins of victory, home court advantage, and the quality of opponents.

Specifically, teams are rated as being better when they beat higher quality opponents, win games by larger margins, and beat teams without having home court advantage. Method E predicts that the team with the better aggregate team metrics will win each game.

Method R:

Method R it is based entirely on NBA player projections (i.e. it doesn't account for wins, losses, and other team metrics). These player projections estimate each player's future performance based on the trajectory of similar NBA players.

A given team's rating is generated by aggregating the ratings of the players on that team with up-to-date depth charts (accounting for injuries, trades, etc.). Method R predicts that the team with the better aggregate rating of its players will win each game.

Then, the participants rated each algorithm using the following replicativeness questions (1 – Strongly disagree; 5 – Strongly agree):

1. This method shares similarities with the process that determines the outcome of NBA basketball games.
2. The way that a team wins a matchup in the NBA is well represented by this method.

We found that participants rated Method E (the team-based method) as more replicative than Method R ($M_{MethodE} = 4.02$ vs $M_{MethodR} = 3.16$, $t(198) = 8.23$, $p < 0.001$), consistent with our expectations.

APPENDIX C: Additional Materials for Study 2c

Predictors	Linear Regression with Standard Error Clustered by Participants (SE)	Linear Regression with Standard Error Clustered by Participants and Algorithm Fixed Effects (SE)	Linear Regression with Random Intercepts by Participants (SE)	Linear Regression with Random Intercepts by Participants and Algorithm Fixed Effects (SE)
Replicativeness	0.32***	0.31***	0.30***	0.28***
	(0.05)	(0.05)	(0.04)	(0.04)
Expensive	-.040	-.059	-.011	-.042
	(.06)	(.06)	(.05)	(.05)
Understanding	.015	.013	.03	.02
	(.05)	(.05)	(.04)	(.04)
Unique	.23***	.22***	.17***	.15***
	(.05)	(.05)	(.04)	(.04)
Intercept	2.64***	2.58***	2.70***	2.72***
	(.24)	(.24)	(.21)	(.21)
df	796	789	796	789

Note: *p<0.1; **p<0.05; ***p<0.01

APPENDIX D: Supplemental Studies

STUDY S1

Study S1 aims to investigate whether participants understood that their chosen prediction algorithm performed worse. In this study, participants had to answer what prediction (1-7) offers the highest payoff on average in addition to making a choice. This will allow us to observe whether any participants who recognize that the replicative algorithm performs worse choose it anyway.

Participants and Procedures. We pre-registered that we would recruit 400 for the study. Four hundred and eighty-five MTurk workers responded to the surveys, 402 of whom finished them and passed the attention checks. Participants' average age was 37 (range: 20-77), and 45% were females.

Study S1 used the same prediction task and incentive scheme as Study 1a. Participants chose between the Optimal Algorithm (i.e., an algorithm that predicts "4") and the Replicative Algorithm (i.e., an algorithm that generates a random number between 1-7) to make a prediction on a 7-sided die roll. They were randomly assigned to two order conditions: "comprehension first" or "choice first." Participants in the comprehension-first group first answered a comprehension question that asked which bet (numbers 1-7) would result in the highest payoff on average before. Participants in the choice-first group chose between the two algorithms before answering the comprehension check question. Finally, all participants answered an exploratory question asking the reason for their choice with an open text box.

Results and Discussion. First, we note that the number of people selecting the correct answer (62.7%, 252/402) was substantially larger than random guessing (14%), suggesting many

or most participants understood what the optimal answer was. There was no significant difference in comprehension accuracy between the two conditions (64.0% in choice-fist vs 61.4% in comprehension-first, $\chi^2(1) = 0.19$, $p = .66$).

Next, we turn to participants who chose between algorithms before answering the comprehension question, as we believe that this condition better replicates Study 1a. Overall, 48.5% (97/200) of participants who made their choice before answering the comprehension question chose the more replicative algorithm. We interpret this as a positive replication of Study 1a, as we believe a substantial present of participants chose the more replicative algorithm even though it offered worse performance. 64% (128/200) of the participants in this condition answered the comprehension question correctly, reporting that a prediction of "4" (that the Optimal Algorithm always makes) results in the highest payoff on average, indicating that many participants who were capable of understanding this performance advantage still chose the more replicative algorithm. Of these participants who answered the comprehension question correctly, 39.8% (51/128) chose the more replicative algorithm. Thus, participants were less likely to choose the replicative algorithm when they were capable of understanding that it performs worse than the optimal algorithm (39.8% vs 63.9%, $\chi^2(1) = 9.73$, $p = 0.002$), but many participants who were capable of understanding this performance advantage still chose the more replicative algorithm. This result suggests that preferences for replicative prediction algorithms may not be driven solely by expectations of higher performance.

Participants who answered the comprehension question before making their choice chose the replicative algorithm directionally less often (40.6%, 82/202) than participants in the choice first condition (48.5%, 97/200, $\chi^2(1) = 2.23$, $p = 0.14$). 61.4% (124/202) of the participants in the comprehension first condition answered the comprehension question correctly, and of these

participants, 35.7% chose the more replicative algorithm. Thus, once again, participants were less likely to choose the replicative algorithm when they were capable of understanding that it performs worse than the optimal algorithm (59.3% vs 35.7%, $\chi^2(1) = 20.29$, $p < 0.001$), but many participants who were capable of understanding this performance advantage still chose the more replicative algorithm. We interpret the 35.7% of participants who chose the replicative algorithm after reporting that the Optimal algorithm's answer of "4" maximizes earnings as expressing a particularly strong preference, as they chose the Replicative algorithm after explicitly acknowledging that the Optimal algorithm performs better.

STUDY S2

In Study S2, we replicate the preference for replicative algorithms in another consumer domain: evaluations of navigation algorithms.

Participants and Procedures. We pre-registered that we would recruit 200 participants from MTurk. Two hundred and forty-three MTurk workers responded to the survey, 202 of whom finished it and passed the attention check. The average ages in the final samples were 36.0 (range 20-73), and 38.3%, were females.

Participants evaluated two prediction algorithms to forecast the driving time from a specific location in the city to a specific destination in the suburbs. The Replicative Algorithm "finds a past trip that has the same route, done at the same time of the day, matches your trip as closely as possible (weather of the day, etc.), and reports the duration of that trip as its estimate". The Average Algorithm "calculates the average time taken to drive the same distance from the city to the suburbs at a similar time of day". Participants rated their likelihood of using each algorithm on a 7-point likert scale (1= *Extremely Unlikely* and 7 = *Extremely Likely*), and the order of the algorithms was counterbalanced. They then explained their rating of each algorithm using an open text box and entered their age and sex to complete the study. Although the Replicative Algorithm more precisely matches the car trip in question, it also relies on only one observation to make its prediction (similar to the replicative algorithm in Study 2b), and we conjecture that it is likely to perform worse than the Average Algorithm as a result.

Results. On average, participants rated the Replicative Algorithm significantly higher than the Average Algorithm (5.56 vs. 5.13, Cohen's $d = 0.18$, $t(200) = 2.59$, $p = .010$). Further, 51.7% of the participants rated the Replicative Algorithm higher than the Average Algorithm, while only 31.8% rated the Average Algorithm higher than the Replicative Algorithm. Thus, the

majority of participants preferred the more replicative prediction algorithm, despite its reliance on a single past trip.

STUDY S3

In Study S3, we test the preference for simulative algorithms using a different scenario: predicting the outcome of the spin of a roulette wheel.

Participants and Procedures. We planned to recruit 300 participants. Three hundred fifty-four subjects responded to the survey, 301 of whom finished it and passed the attention check. The average age in the final sample was 36 (range: 20-77), and 51.5% were females.

Participants imagined they were predicting the outcome of a roulette-wheel spin with 37 pockets labeled 0, 1, 2, …, 36. They also imagined their reward for the prediction would increase linearly with their prediction accuracy ($37 if they predicted perfectly, a reduction of $1 for each unit of error). They viewed three algorithms and rated their likelihood of using each algorithm in predicting the outcome of the spin on a 5-point Likert scale. The Optimal Algorithm always predicts 18, and the Replicative Algorithm spins the roulette wheel and submits the number of the pocket that the ball lands in as its prediction. The Simulation Algorithm spins the roulette wheel 10 times and submits the number that would earn the most over those 10 spins as its prediction. We only include one simulative algorithm because we did not observe significant differences between the two different simulative algorithms in Study 4.

Our hypothesis suggests people will prefer the Replicative Algorithm and Simulation Algorithm to the Optimal Algorithm because both replicate the process of spinning a roulette wheel. However, the Optimal Algorithm maximizes participants' expected earnings because 18 is the median number on the roulette wheel. Although the Simulation Algorithm does not always predict the optimal number (18), it does provide answers that are closer to optimal on average than the Replicative Algorithm. Thus, people who pick the Simulation Algorithm are much better off than those who pick the Replicative Algorithm.

Results and Discussion. Replicating the finding that people prefer replicative algorithms, we found participants rated the Replicative Algorithm higher than the Optimal Algorithm (M=3.04 vs 2.65, Cohen's $d = 0.29$, $t(300) = 3.74$, $p < .001$). Moreover, as predicted, they preferred the Simulation Algorithm to the Optimal Algorithm ($M = 3.45$ vs. 2.65, Cohen's $d = 0.63$, $t(300) = 7.94$, $p < .001$). Further, participants also preferred the Simulation Algorithm to the Replicative Algorithm ($M = 3.45$ vs. 3.04, Cohen's $d = 0.36$, $t(300) = 5.81$, $p < .001$), suggesting that simulation algorithms may be attractive in addition to outperforming replicative algorithms that rely on fewer trials.

APPENDIX E: Additional Materials for Study 3

Example of the feedback participants got after making a choice

The result of the dice roll is 5.

Method T's prediction is 3.

Method V's prediction is 4.

[If participants were in the target-more-accurate condition, they read:]

You choose Method T, yielding a $0.05 bonus.

[If participants were in the same performance condition, they read:]

You choose Method T, yielding a $0 bonus.

REGRESSIONS FOR STUDY 3

Predictors	Linear Regression with Standard Error Clustered by Participants (SE)	Linear Regression with Standard Error Clustered by Participants and 3-way interaction (SE)	Linear Regression with Random Intercepts by Participants (SE)
Choice Number	-.009*	-.010	-.016**
	(.005)	(.005)	(.006)
More Accurate	.367**	.359**	.849**
	(.144)	(.144)	(.28)
More Replicative	.379***	.371**	.728**
	(.141)	(.141)	(.28)
Choice Number * More Accurate	.016***	.016**	.028***
	(.006)	(.008)	(.007)
Choice Number * More Replicative	.002	.003	.003
	(.006)	(.009)	(.007)
More Accurate * More Replicative	-0.004	0.012	.005
	(.199)	(.22)	(.38)
Choice Number * More Accurate * More Replicative	N/A	-.0016	N/A
		(.012)	
Intercept	.021	.025	-.0087
	(.10)	(.10)	(.20)
df	795	794	794

Note: *p<0.1; **p<0.05; ***p<0.01