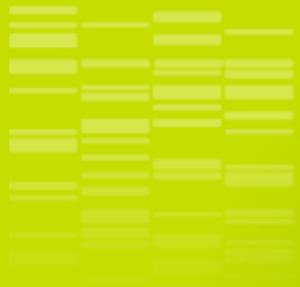


Food labeling for human beings

The role of cognitive limits in GDA vs. TL debate

Crosetto - Muller - NIBS – Nottingham, 22 april 2015

This talk

What : labels.

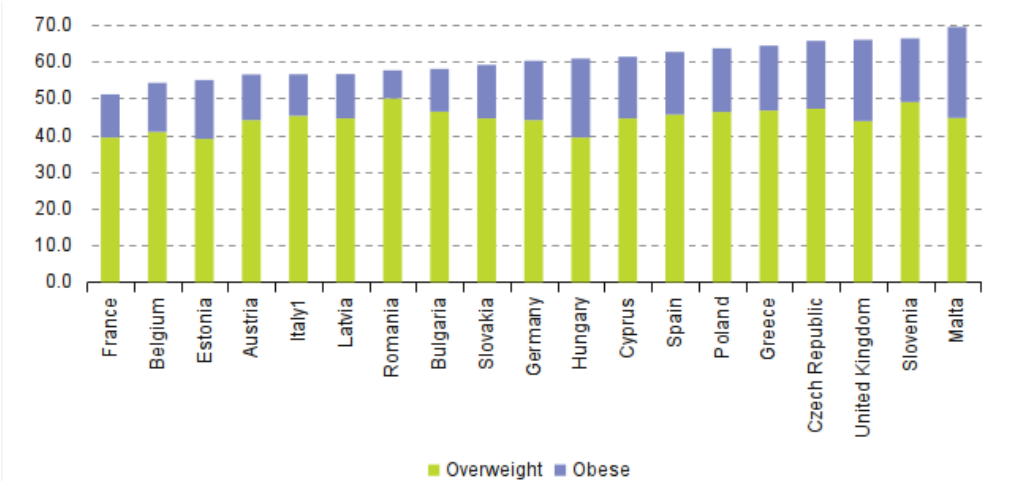
- Key issue in obesity prevention
- Lots of applied research
- A heated political debate

How : Incentivized lab experiments

- Controlled setting
- Incentivized choices
- A representative sample of subjects
- Artificial environment $\neq$ real shopping / eating habits
- Aim : assessing performance of different labeling schemes *per se*

Motivation

- More than one-third (34.9%) of U.S. adults are obese.
- For the UK, this share is 22.1% for adult men, 23.9% for women
- EHIS data, Eurostat 2009 :



- Obesity $\Rightarrow$ higher incidence of hearth diseases, diabetes,...
- The estimated annual medical cost of obesity in the U.S. was $147 billion in 2008 U.S. dollars

GDA vs. TL : a long debate

We focus here on three possible labeling formats

Guideline Daily Amount

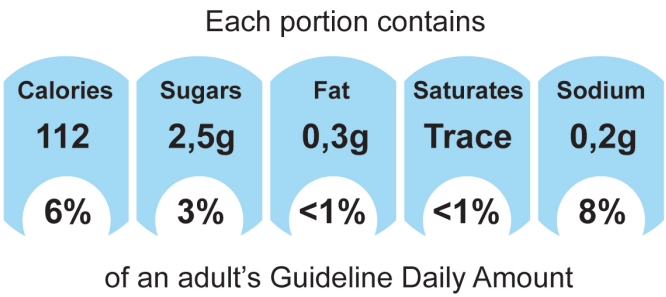
- Nutrition information is expressed as a % of the GDA.
- It can be mono- or multi-dimensional.
- Gives information + a slight suggestion.
- Relies on the ability of the consumer to process the information...
- ...keep track of things bought, contrast & compare.

TL

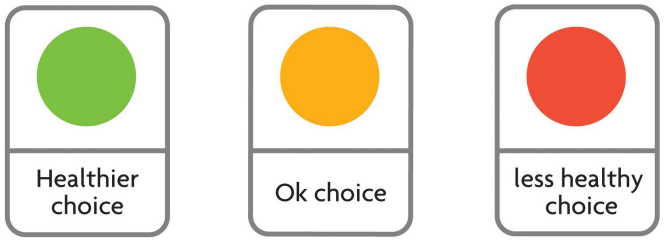
- Nutrition information is expressed as *color-codes*.
- Three levels : red, amber, green.
- Green : good. Red : bad.
- Less informative, more salient information.

GDA, TL, GDA+TL

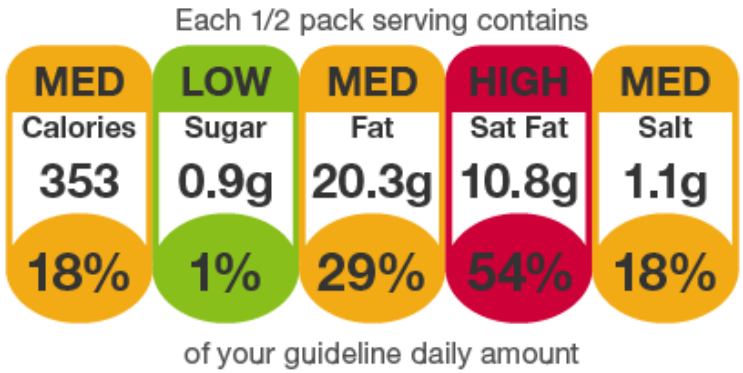
Guideline Daily Amount (GDA)



Traffic Lights (TL)

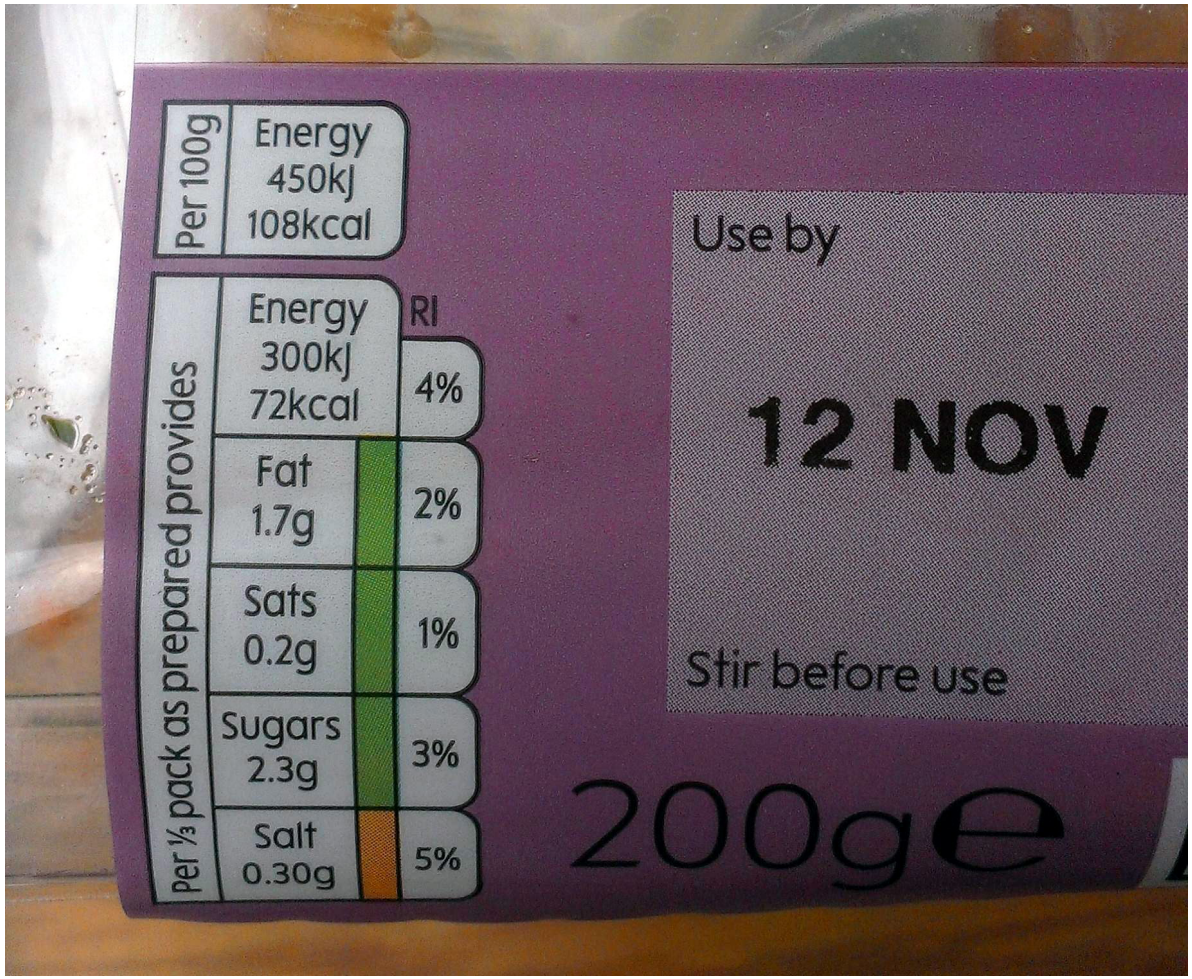


GDA+TL



English TL, live !

Waitrose mixed beans salad, King's Cross



Some experimental evidence, I

Kelly et al. 2009

- 790 interviews in Australian grocery stores
- GDA, color-coded GDA, TL, TL + overall rating
- Each consumer exposed to two *fake* products
- Basic questions to assess performance :
  1. level of nutrient ? "a lot", "moderate amount", "small amount"
  2. which of the two is the healthier product ?

The question seems to favour TL...

Results

- TL better than GDA in both tasks (correct : 78% TL+, 64% GDA)

Some experimental evidence, II

Borgmeier and Westenhofer, 2009

- Randomised experiment with 420 subjects in Germany
- "healthy choice", GDA, color-coded GDA, TL
- Each consumer exposed to couples of *real* products
- Basic questions to assess performance :
  1. which of the two is the healthier product ?
  2. suppose you had to make your shopping for one day...

Question 1 favours TL ; question 2 depends on preferences...

Results

- TL better than GDA in comparison (24.8/28 TL+, 22.8/28 GDA)
- No significant difference in shopping task

Some experimental evidence, III

FSA, 2005

- Questionnaire over 2676 subjects
- GDA, TL, GDA+TL
- Two conditions evaluate, for two key nutrients :
  1. is the product high, medium or low in each of the nutrients
  2. which product has the highest amount of nutrient x, y

Question 1 favours TL ; question 2 favours GDA.

Results

- TL better than GDA in question 1
- GDA better than TL in question 2

It is the question, stupid !

In the existing literature, the questions seem to determine the answers

Papers that ask for relative ranking favor TL

- ...especially if asking for three categories
- ...especially within a limited range of products (2,3)
- ...especially if the label is monodimensional

Papers that ask for absolute ranking favor GDA

- ...especially if asking for how much more x is in product y
- ...especially if asking on more than one dimension
- ...especially for better educated or richer people

Are we asking the right questions ?

Our experiment, I

Question : what label is better to *build a healthy diet* ?

What we ask the subjects to do

We tell the subject he has been hired as a *nutritionist* in a public canteen, that caters to all sorts of people for the whole day. Subjects must compose a *daily menu* for the canteen. Subjects are paid *if and only if* the chosen menu satisfies a set of pre-determined nutritional criteria. To guide subjects in their choices information might be provided, in the form of TL, GDA, or both.

Incentives

- If the daily diet built satisfies nutritional constraints $\Rightarrow$ flat fee
- Several daily diets to build

Our experiment, II

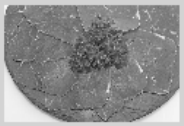
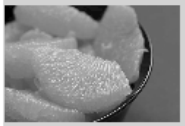
Features of the task

- No preferences $\Rightarrow$ *incentivized*, *third-person task*
- What is a diet ? vague $\Rightarrow$ *focus on daily consumption*
- Question biased by construction $\Rightarrow$ *towards GDA*
- "Realistic" task $\Rightarrow$ gets near to what we want the label to do.

Our task : screenshot

Petit dejeuner		Lait frais entier		Nectar de fruits exotiques		Eau gazeuse		Lait aromatisé
		Céréale type All Bran		Pain de mie		Orange		Pain suédois
Dejeuner		Salade frisée		Pomme de terre à l'huile		Viande des Grisons		Pâté de foie de volaille
		Sandwich crudités fromage		Sandwich crudités rosbif		Pot au feu		Sandwich type libanais (fallafel)
		Pamplemousse frais		Fraise		Mousse de fruit		Salade de fruits
Collation		Gâteau de Savoie		Petit-suisse		Meringue		Kiwi
		Avocat vinaigrette		Laitue		Asperge grande		Rollmops de hareng
		Flageolet		Pâtes complètes		Quinoa		Châtaigne
Diner		Haricot rouge		Pâtes fraîches		Poivron rouge grillé		Carotte

Our task : screenshot

Petit dejeuner		Lait frais entier		**Nectar de fruits exotiques**		Eau gazeuse		Lait aromatisé
		**Céréale type All Bran**		Pain de mie		Orange		Pain suédois
Dejeuner		**Salade frisée**		Pomme de terre à l'huile		Viande des Grisons		Pâté de foie de volaille
		Sandwich crudités fromage		Sandwich crudités rosbif		Pot au feu		**Sandwich type libanais (fallafel)**
		Pamplemousse frais		Fraise		**Mousse de fruit**		Salade de fruits
Collation		Gâteau de Savoie		Petit-suisse		Meringue		Kiwi
		Avocat vinaigrette		Laitue		Asperge grande		Rollmops de hareng
		Flageolet		**Pâtes complètes**		Quinoa		Châtaigne
Diner		Haricot rouge		Pâtes fraîches		Poivron rouge grillé		Carotte

Our experiment, III

We give the best chances to GDA

- The targets are cardinal numbers
- All information is in one and only one screen (no memory problems)
- No role for preferences
- No time pressure
- Possibility to just do all the needed computations and walk away with the money

...still, what if TL wins ?

- The task is cognitively complex (many choices, lots of numbers)
- The presence of TL might simplify the task (heuristics)
- It is suboptimal for a *homo oeconomicus*...
- ...but maybe not for *homo sapiens*.

Number of dimensions

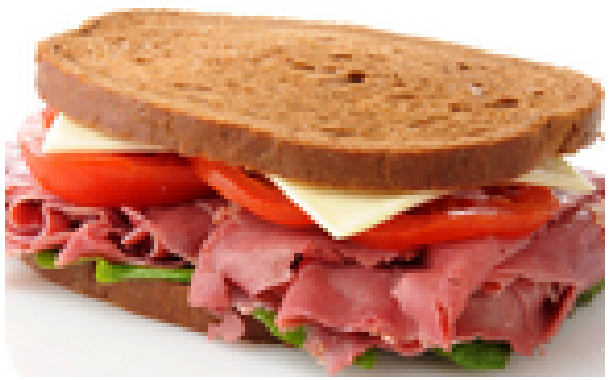
Data are multidimensional. We consider three cases :

- 1-dimension Kcal only are displayed.
- 4-dimension Kcal + 'bad' nutrients : salt, sugar, fat.
- 7-dimension 4d + 'good' nutrients : vitamin C, fiber, calcium.

Example : GDA, 4D

Tarte aux poireaux	
Energie :	12.3
Sucres :	1.7
Graisses :	46.4
Sel :	19.1

Example : TL, 4D



Sandwich crudités rosbif

Energie :	14.4
Sucres :	
Graisses :	
Sel :	

Example : GDA + TL, 6D

Pêche

Energie :	3
Sucres :	14
Graisses :	0
Sel :	0
Vitamines :	11
Fibres :	12
Calcium :	2

Daily diet

A daily **diet** is composed of *twelve* food items over *four* meals :

Daily base		120g bread, 10g butter, 20g oil
Breakfast	*Drink*	The, coffee, milk, hot chocolate, juice...
	Main course	Bread, sweets, *viennoiseries*...
	Fruit	Fruit, jam...
Lunch	*Entrée*	Light dishes, ham, paté...
	Main course	Sandwich, pizza, pasta...
	Seasoning	Oil, butter, spices & herbs
	Dessert	Fruit, sweets...
Afternoon snack	-	Sweets
Dinner	*Entrée*	Light dishes, ham, paté...
	Main course	Meat or fish
	Side	Vegetables, rice...
	Dessert	Fruit, sweets...

Optimal daily diets

With an algorithm we built **optimal** and anti-optimal daily diets satisfying (or not) *jointly* :

- No duplicate products
- 90%GDA $\leq \sum$Kcal $\leq$110%GDA
- $\sum$ Sugar $\leq$100%GDA
- $\sum$ Fat $\leq$100%GDA
- $\sum$ Salt $\leq$100%GDA
- $\sum$ Fiber $\geq$100%GDA
- $\sum$ Vitamin C $\geq$100%GDA
- $\sum$ Calcium $\geq$100%GDA

We selected **12** optimal and **24** anti-optimal *daily diets*

Screen creation

We build screens assembling **4** menus

- 2 optimal menus
- 2 anti-optimal menus
- menus are scrambled and mixed...
- ...and we get a 12 rows $\times$ 4 columns screen.
- In each screen subjects must submit 11 4-way choices.

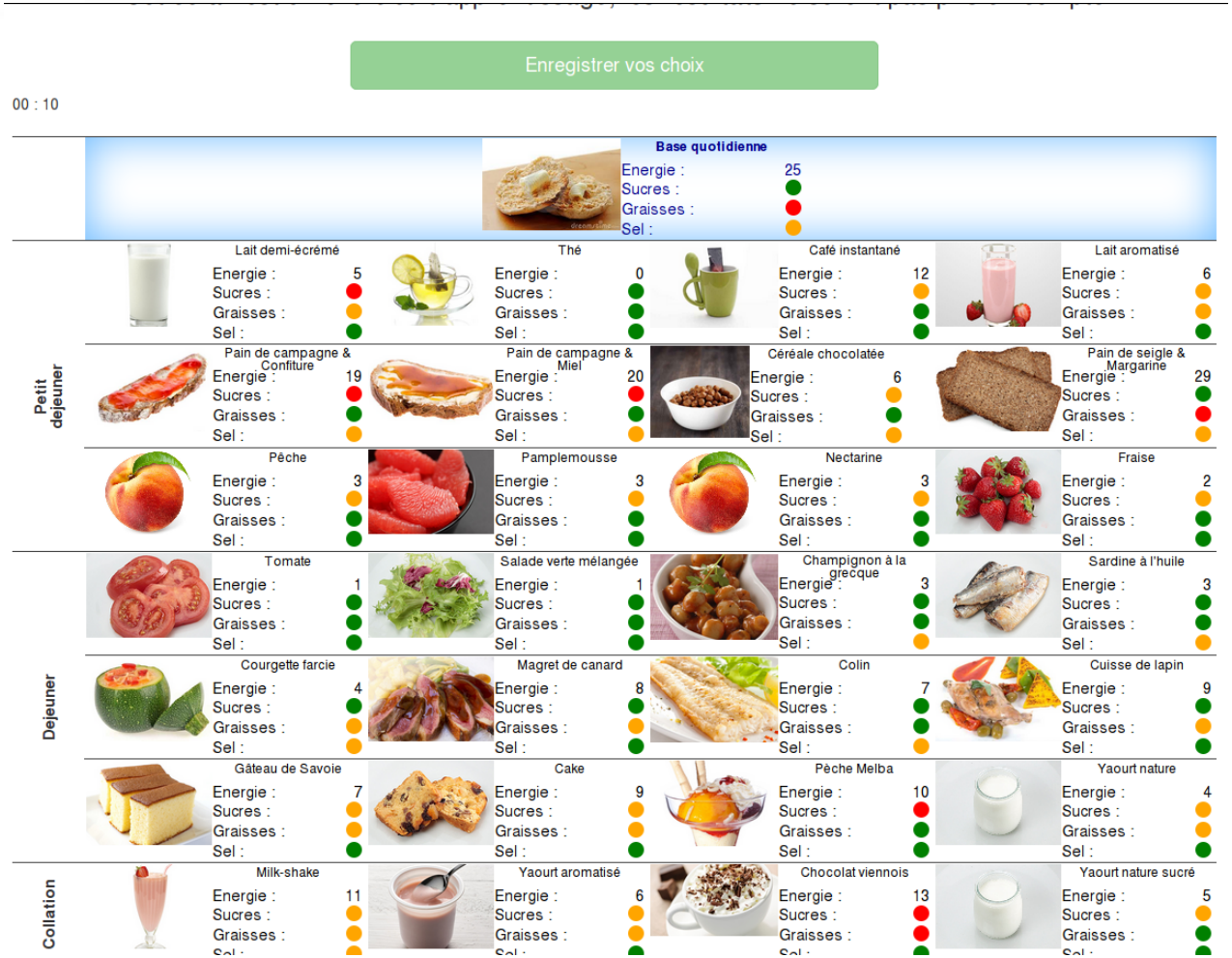
We built 5 screens, each in 4 random orders

Treatment screen : GDA, 4D

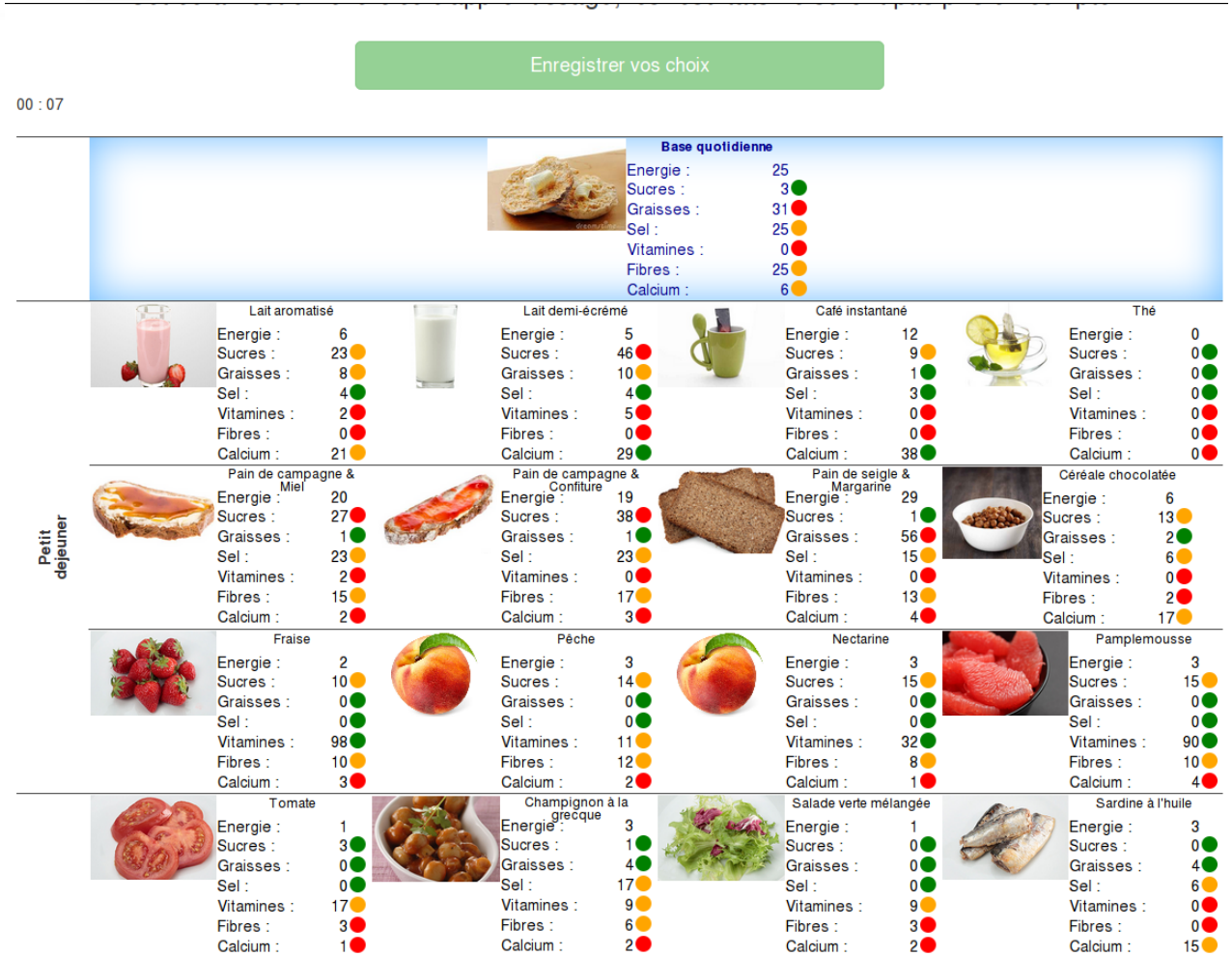
Enregistrer vos choix

00 : 19

Treatment screen : TL, 4D



Treatment screen : GDA + TL, 6D



Random play on a screen

As a benchmark, we simulate random play on our screens

	Likelihood (%) of passing the criterion when playing randomly								
	Kcal90110	*AGS*	*Sucre*	*Sel*	*Fibres*	*Vit*	*Ca*	*4d*	*7d*
S1	29.68	48.18	68.71	72.98	51.70	58.57	47.91	4.61	0.57
S2	32.78	57.84	70.27	58.76	51.00	48.29	40.03	4.78	0.49
S3	30.64	53.75	56.36	67.23	52.50	64.91	45.11	2.54	0.49
S4	31.87	49.03	60.71	67.56	42.14	61.39	43.69	2.98	0.30
S5	29.98	25.85	65.66	49.40	35.98	58.36	24.94	0.86	0.08
Avg	30.99	46.93	64.34	63.19	46.66	58.30	40.34	3.15	0.39

Hard for random players, especially for 4D and 6D

Experiment 1 - treatments

Experiment 1 gives us a benchmark of performances in the cleanest possible environment

We recruit subjects from two different subpopulations :

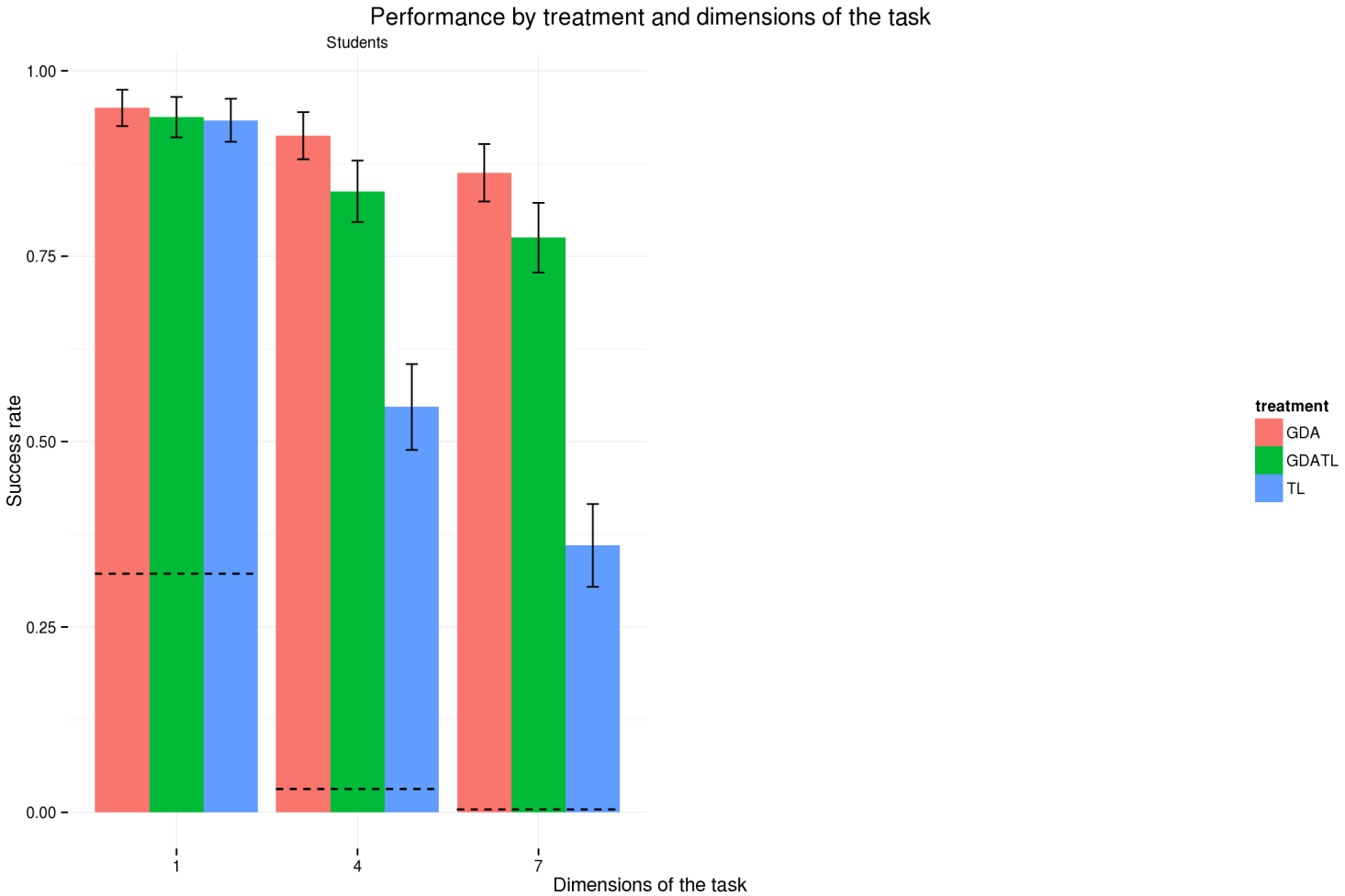
1. Grenoble INP master students $\Rightarrow$ *homines œconomici ?*
2. General population $\Rightarrow$ *homines sapientes ?*
3. 3 sessions (GDA, TL, GDA+TL), 15 subjects/session

Experiment in a pure-between structure

	GDA	**TL**	**GDATL**
Students	16	16	15
General Population	14	12	13

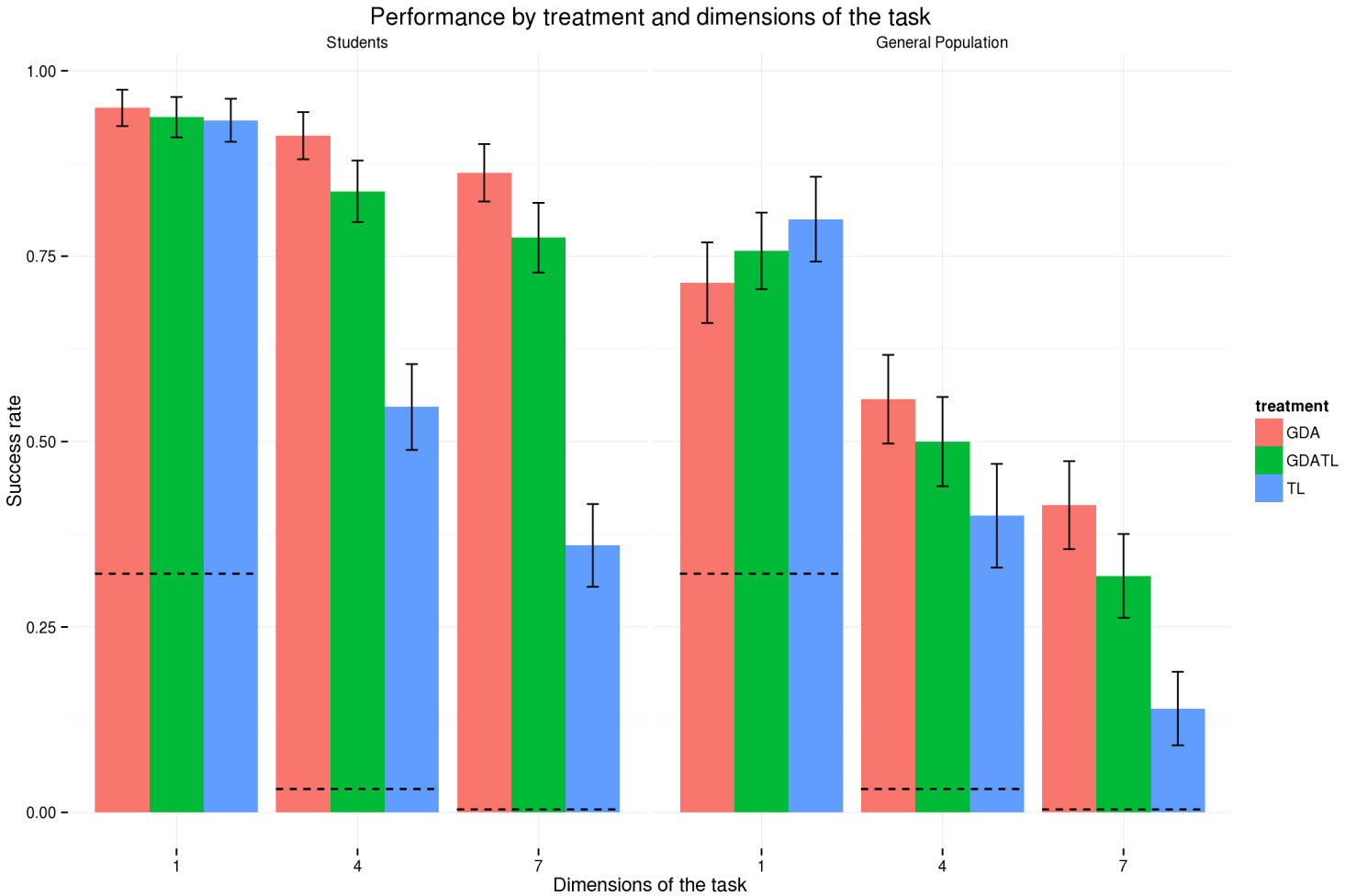
Table : The structure of the experiment and participants

Performance in the tasks



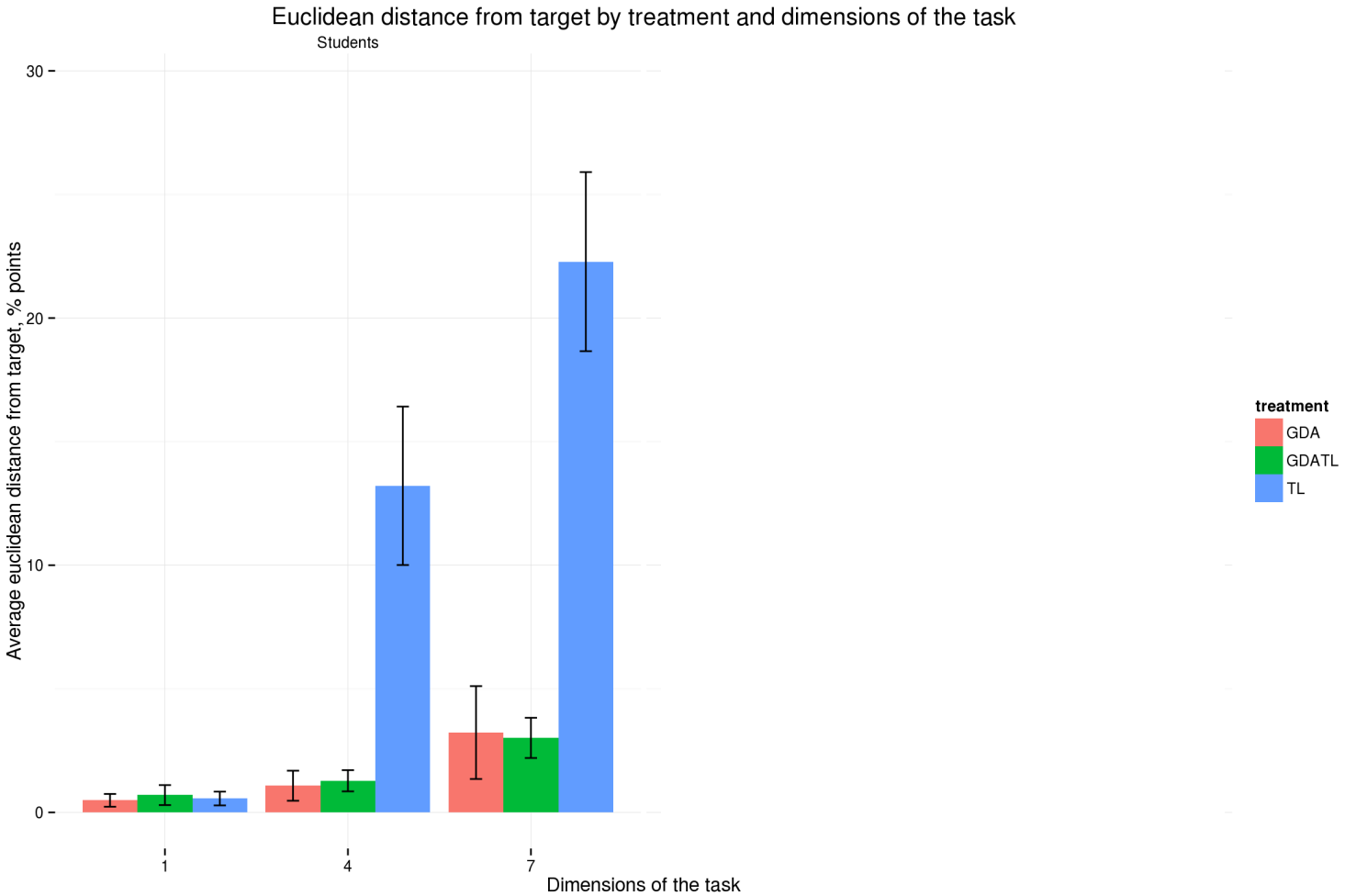
1 = always correct, 0 = always wrong. The dotted line represents random play

Performance in the tasks



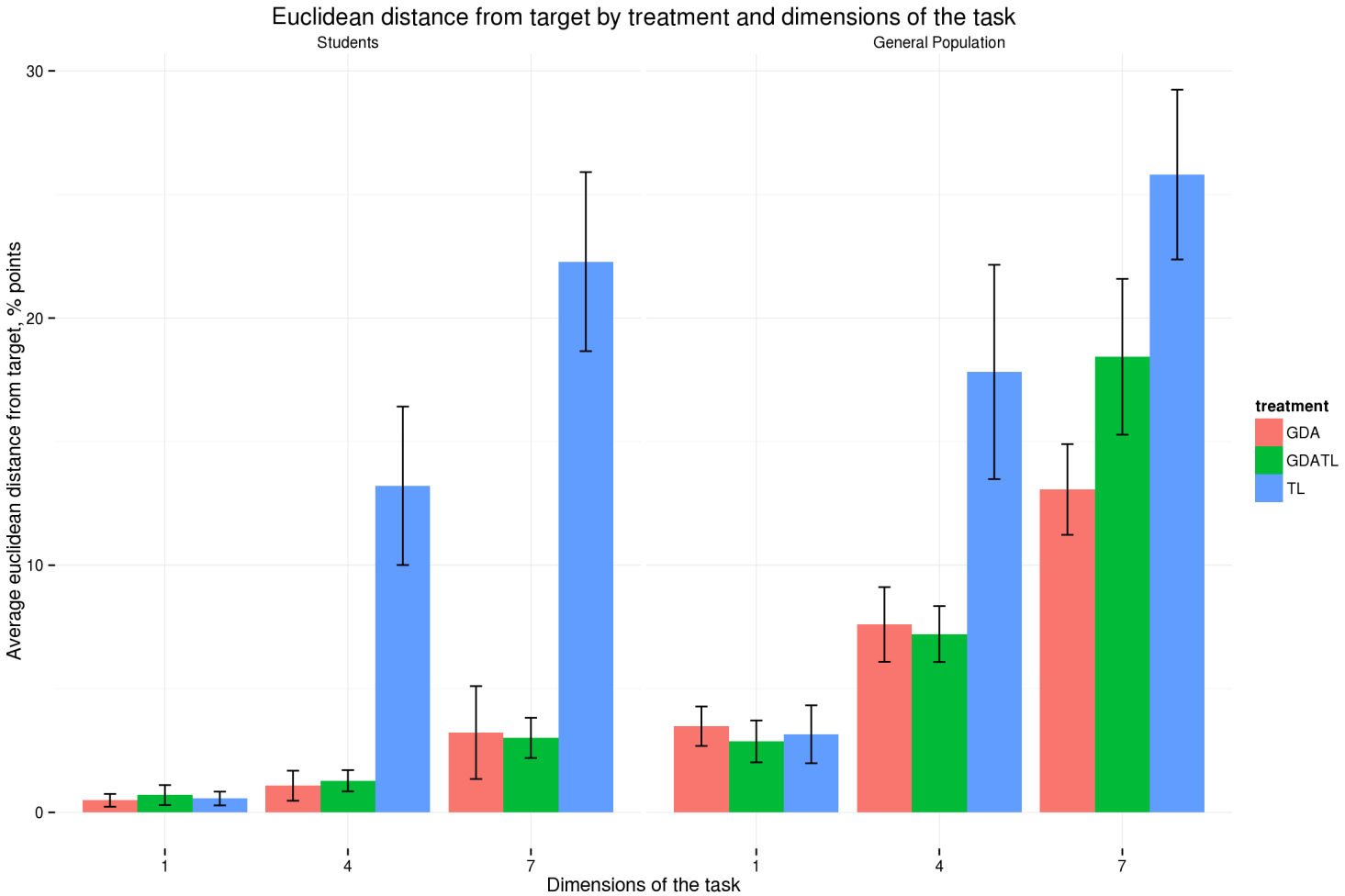
1 = always correct, 0 = always wrong. The dotted line represents random play

Euclidean distance from target(s)



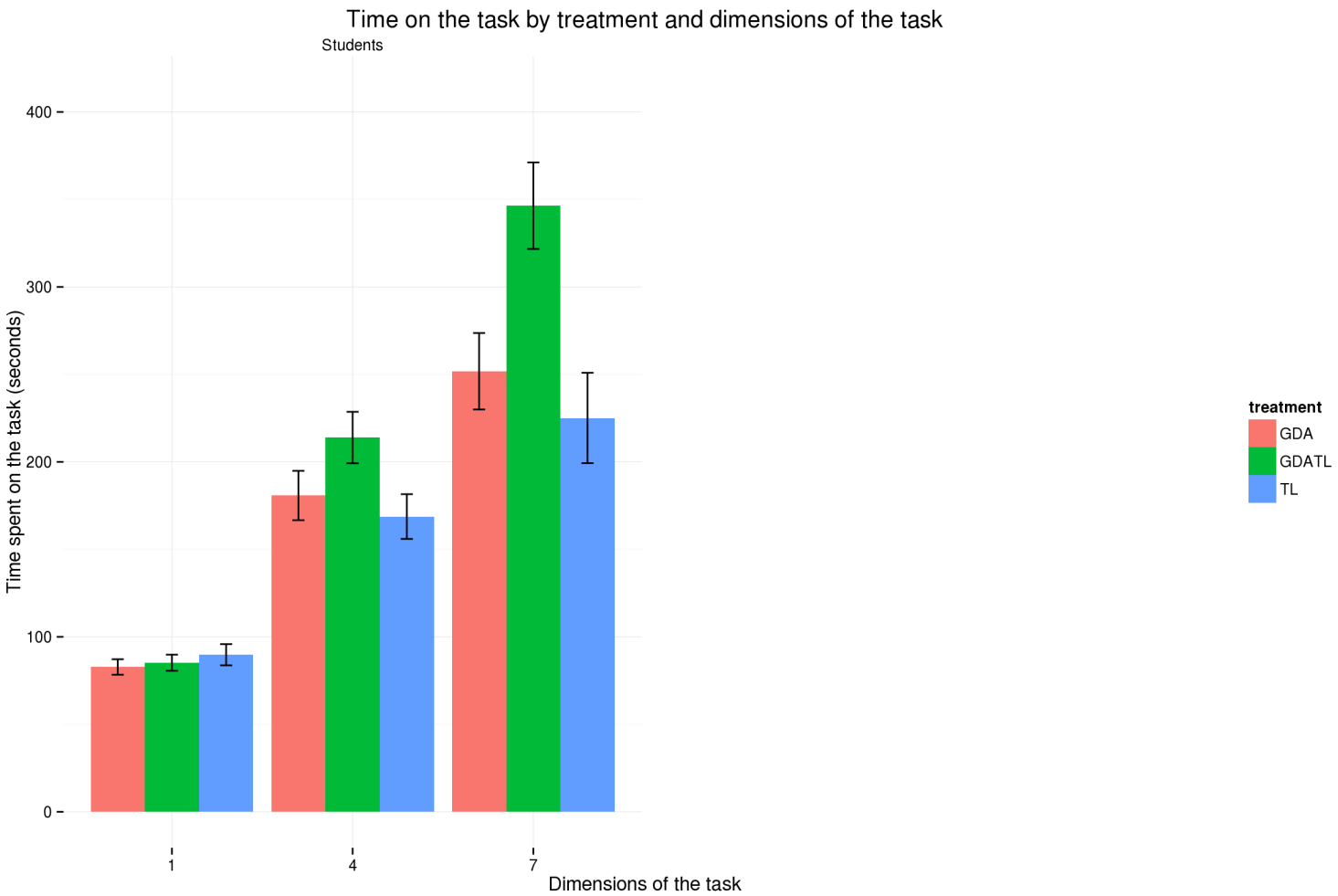
Euclidean distance from target. 0 = on target on all targets.

Euclidean distance from target(s)



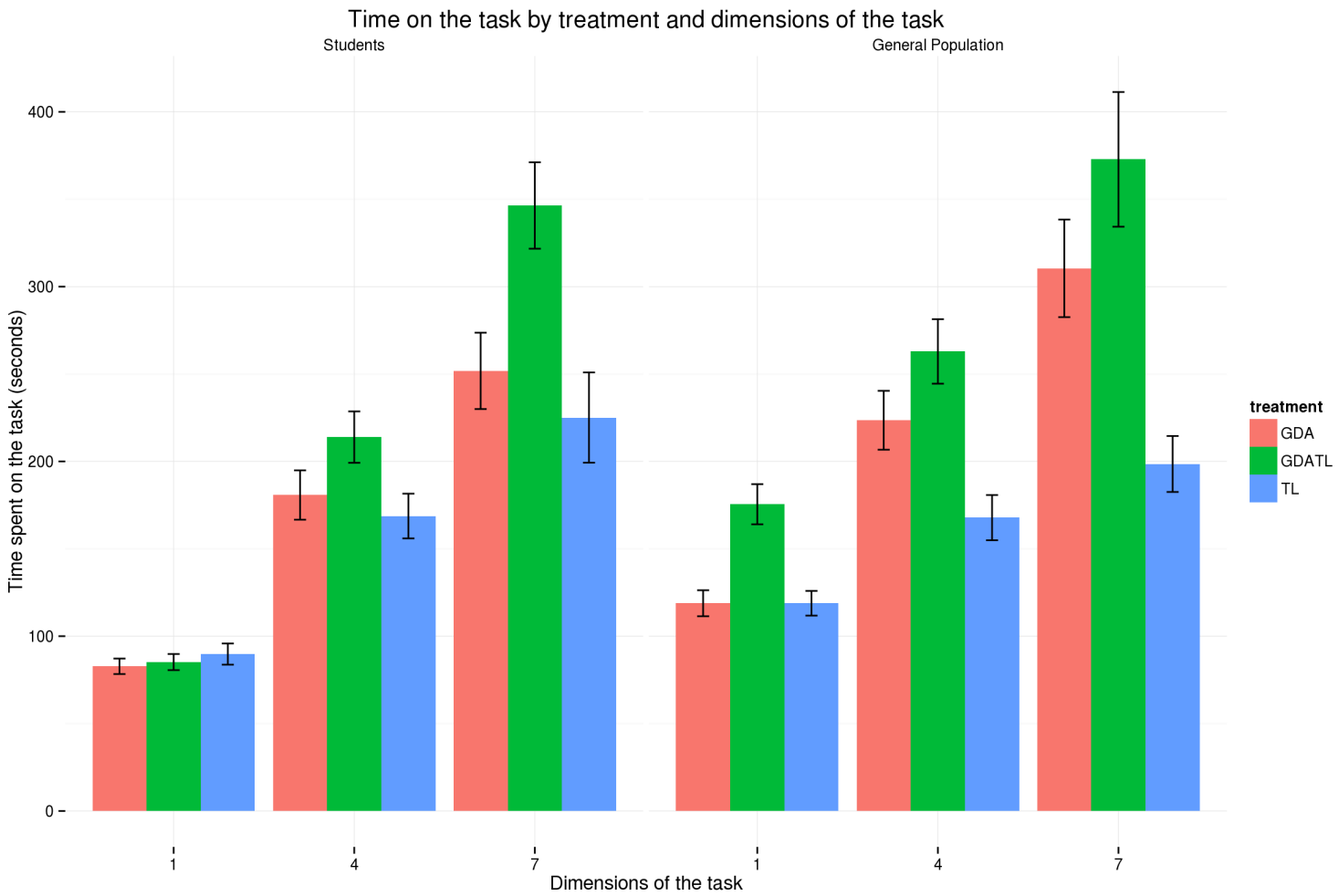
Euclidean distance from target. 0 = on target on all targets.

Time spent on the task



Time in seconds. Error bars are 95% c.i. for the mean

Time spent on the task



Time in seconds. Error bars are 95% c.i. for the mean

General results

$$\text{GDA} \sim \text{GDATL} > \text{TL}$$

Students :

1. tend to do all the computations needed, taking a lot of time
2. kcal-only is identical in the three treatments $\Rightarrow$ identical performance
3. performance does *not* decrease with dimensions, time goes up
4. GDA $\geq$ GDATL $>$ TL

General population :

1. tend to do all the computations needed, taking a lot of time, but fail
2. in TL significantly less time, some switch to heuristic
3. results qualitatively similar to students, at lower level
4. GDATL slower than GDA for $\leq$ performance

(provisional) conclusions

What do we learn ?

- In our artificial environment, GDA $>$ TL. Does it matter ?
- Yes ! upper bound of the possibility of using these tools
- Yes ! differences students / population significant and in the expected direction
- Yes ! we have a benchmark allowing us to introduce 'realistic' features
- ...like time.

Experiment 2 - treatments

Experiment 2 makes a leap to a more realistic setting

- Limited time : 2min per task ; no sheet of paper to aid in computations.
- Extra screens : control for nutritional beliefs, math
- Representative sample, 174 subjects.
- 14 sessions, $\sim$ 15 subjects/session

As before, a pure-between structure

	GDA	TL	GDATL
General population	56	63	55

Table : The structure of the experiment and participants.

Hypotheses

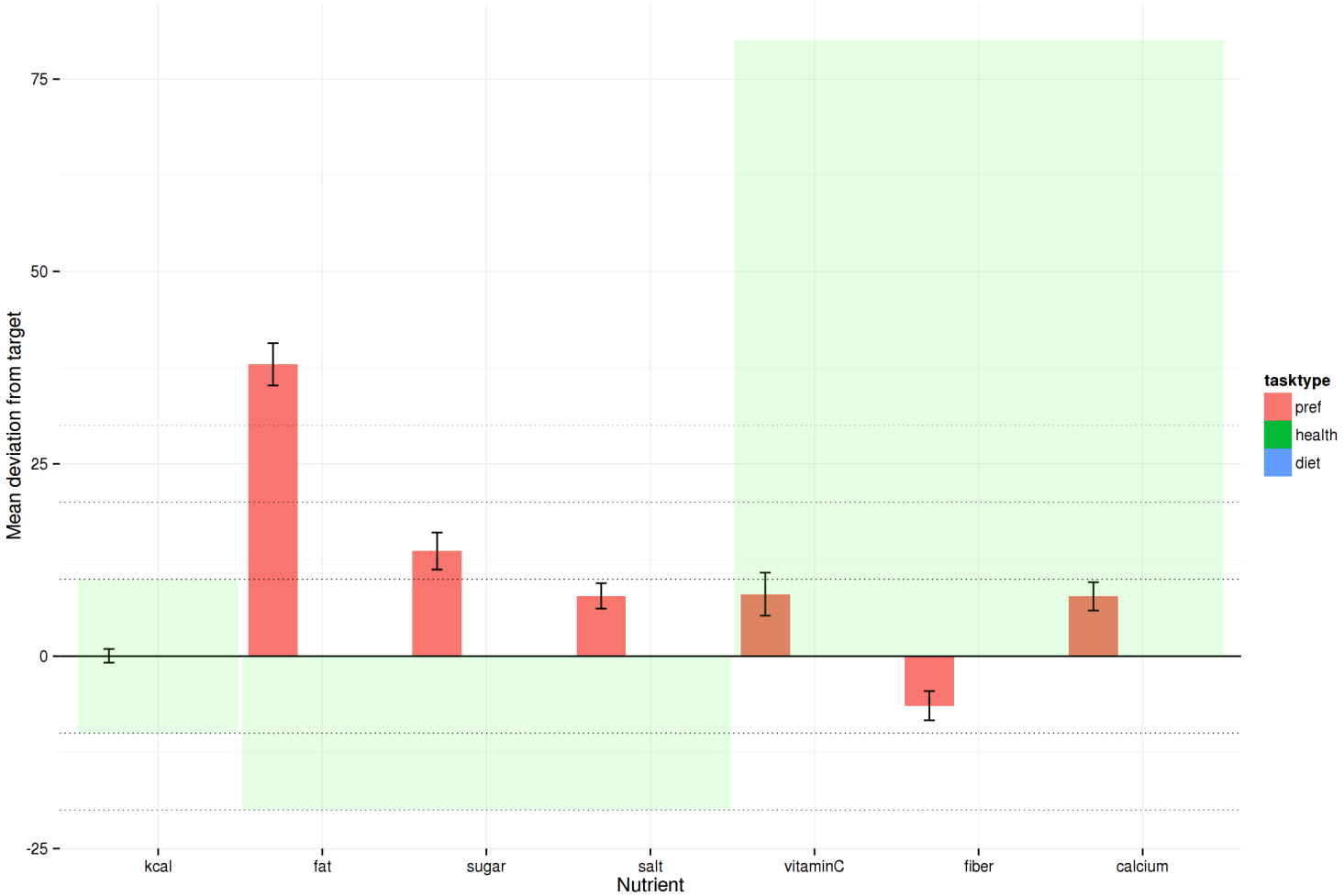
Hypotheses

- The presence of a time limit should induce the use of *fast* heuristics...
- ...leading to a better performance of TL
- ...especially for high-dimensional tasks.

Preferences, Beliefs, Labels

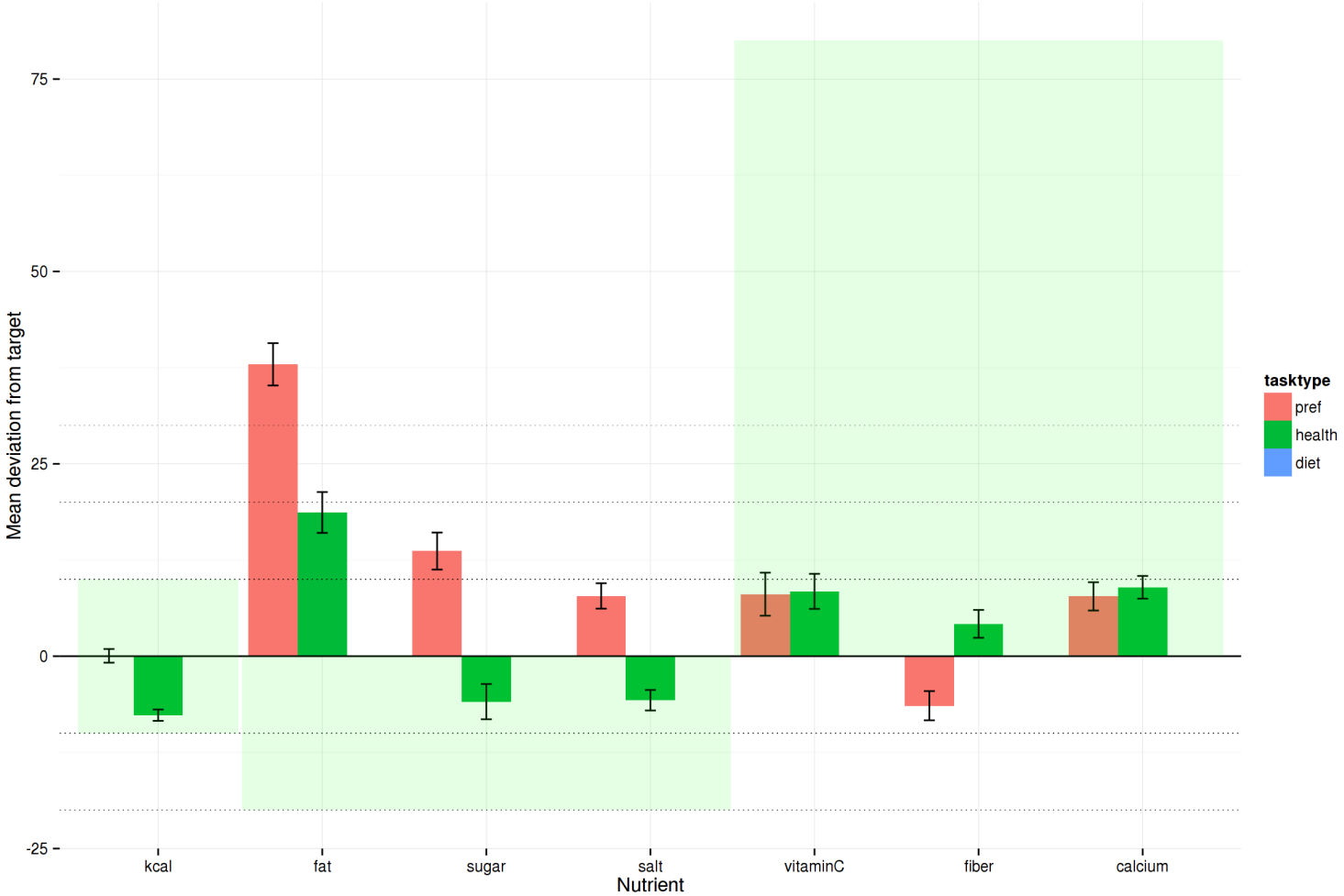
- The new tasks allow us to record subjects' preferences...
- ...and their belief about the healthiness of food items.
- Do label allow subjects to improve on those ?

No labels - Preferences



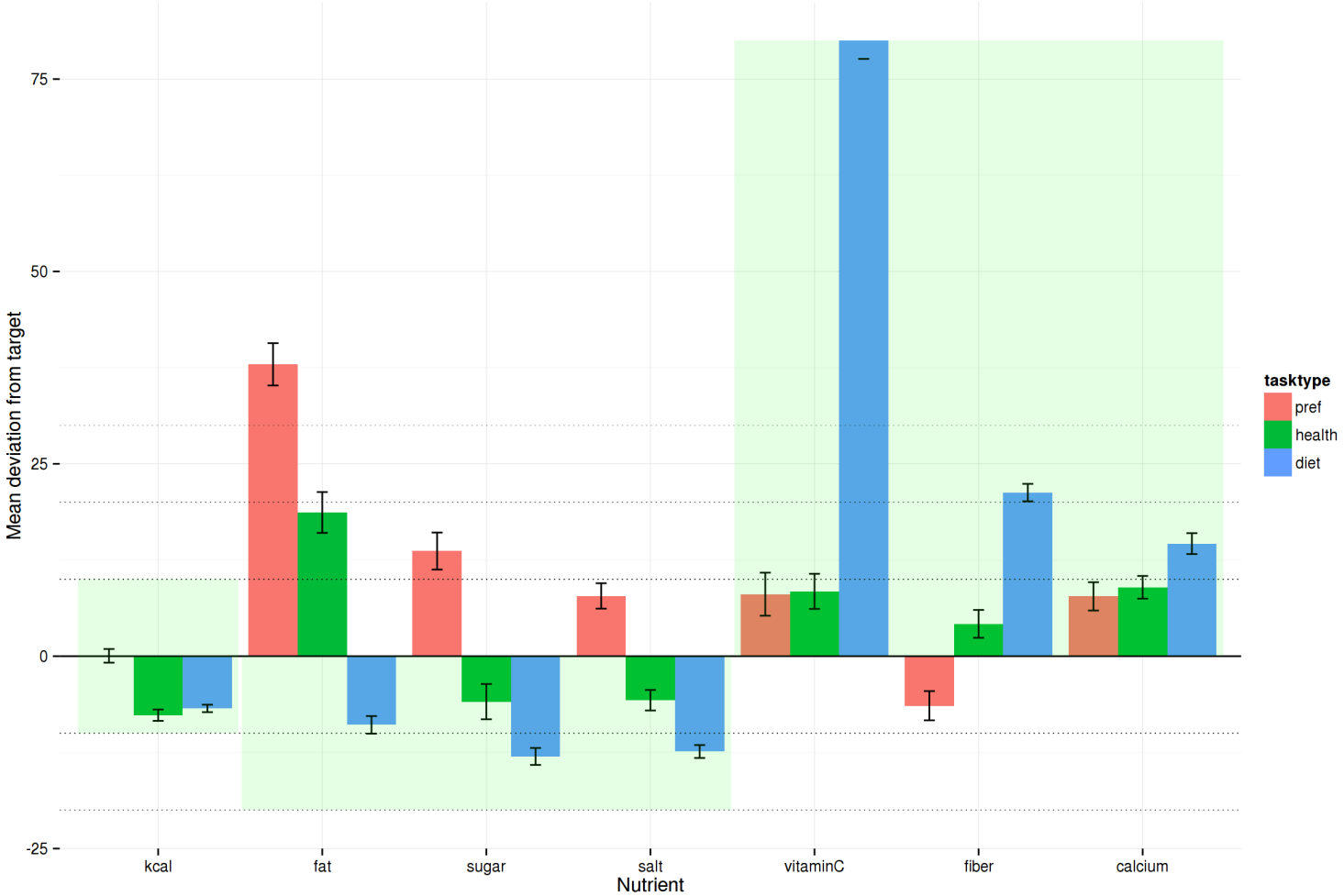
Preferences task, average performance by nutrient.

No labels - Healthiness



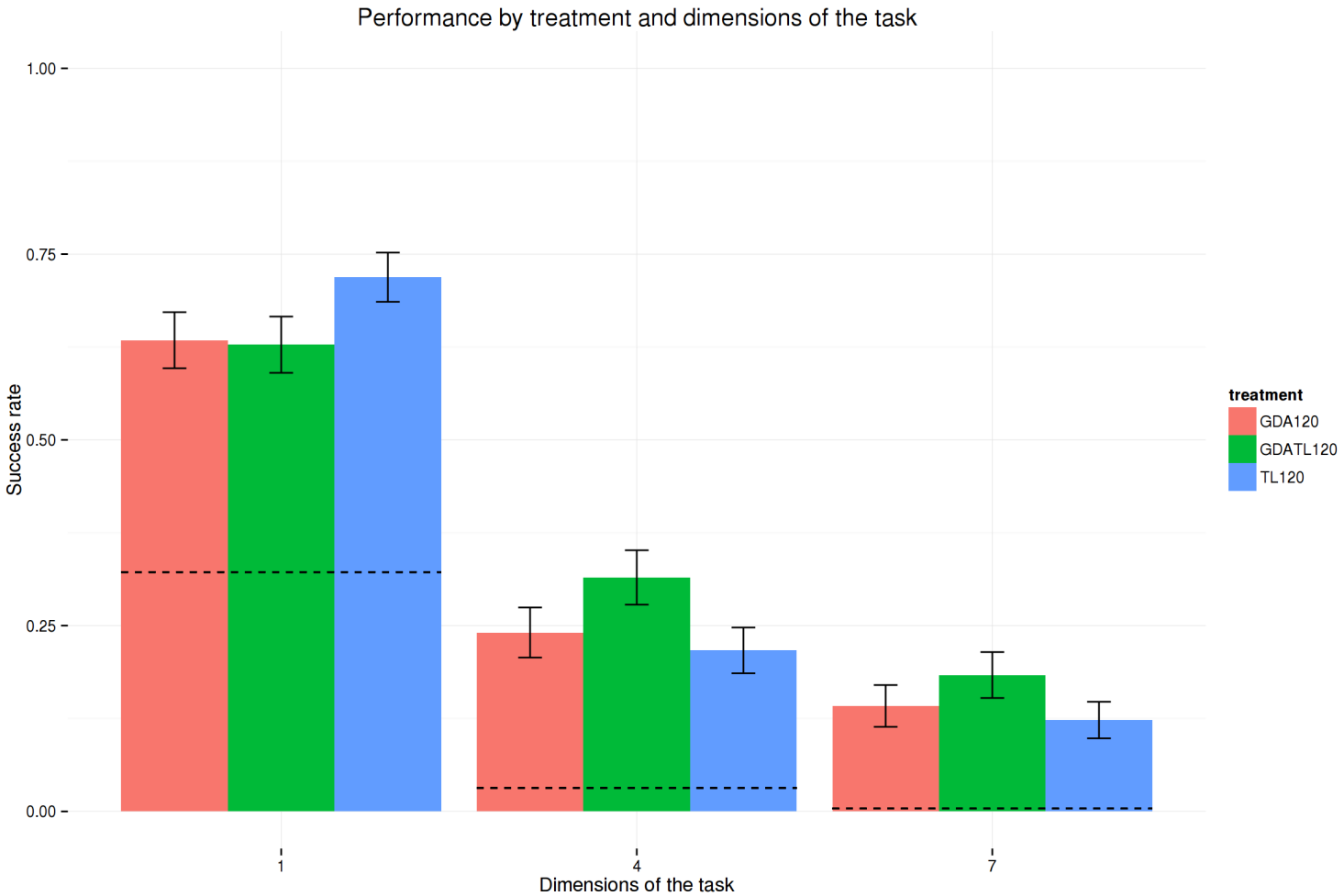
Preferences and healthiness tasks, average performance by nutrient.

Labels (all, mixed)



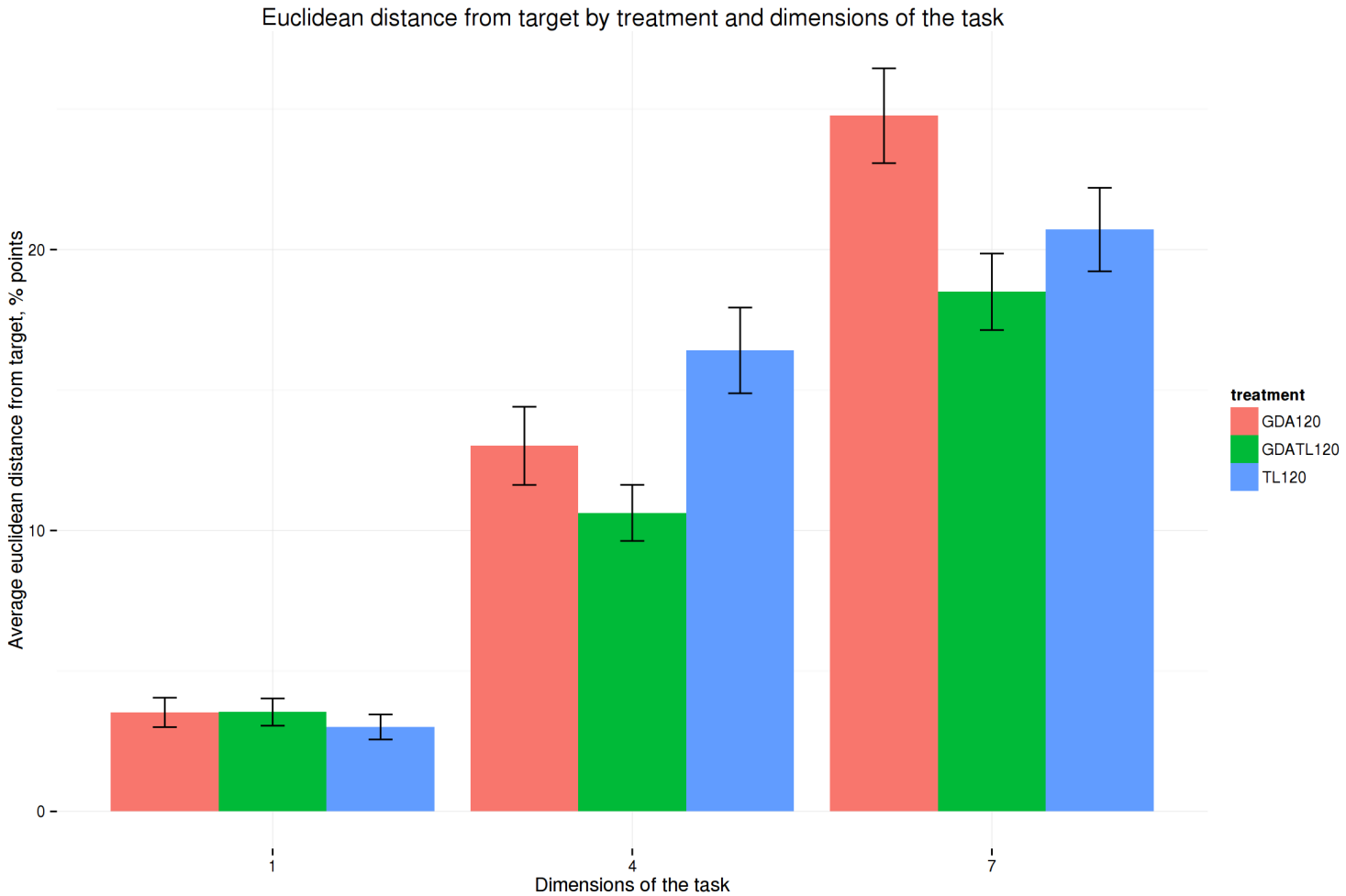
Preferences, health and diet tasks, average performance by nutrient.

Performance in the tasks



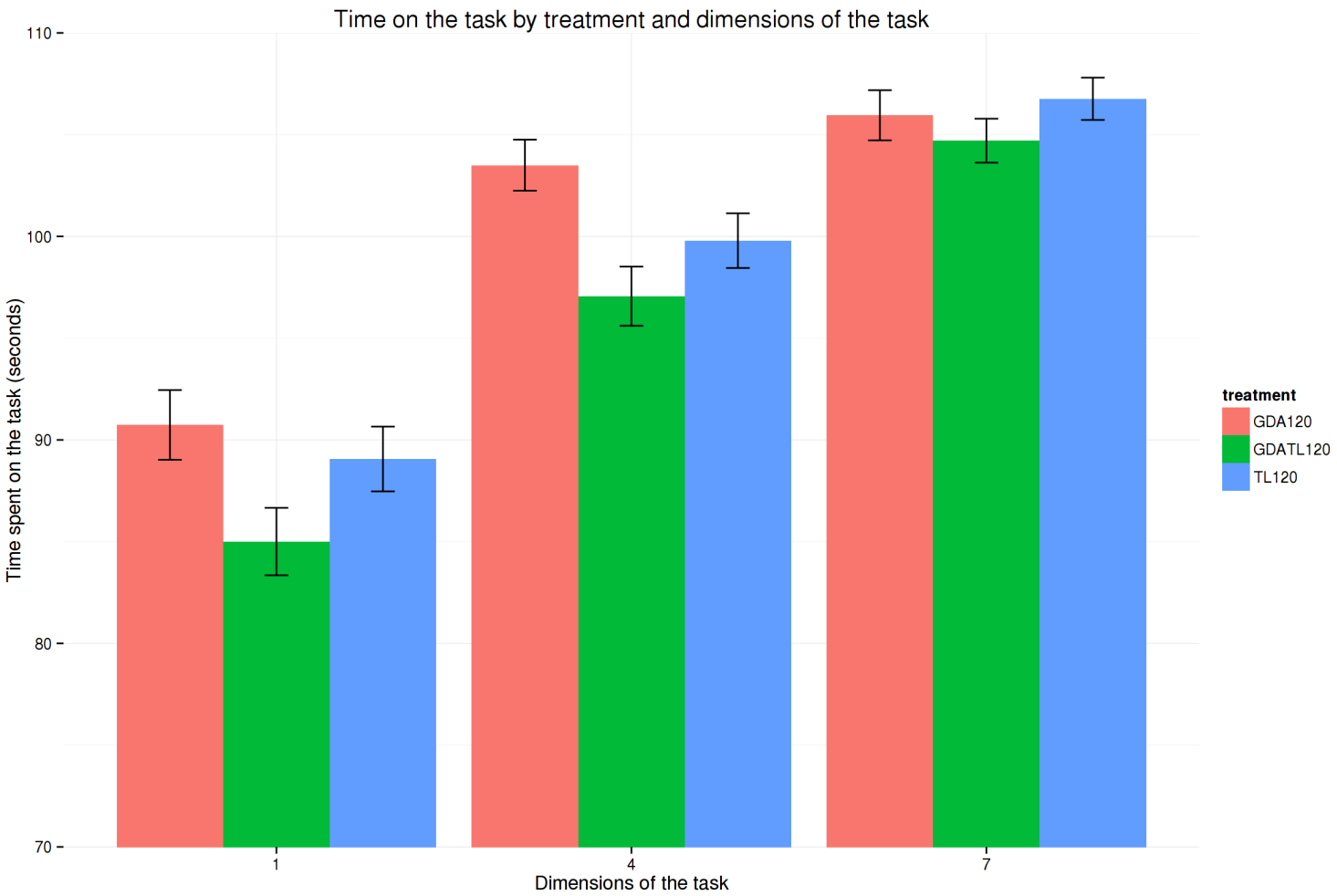
1 = always correct, 0 = always wrong. The dotted line represents random play

Euclidean distance from target(s)



Euclidean distance from target. 0 = on target on all targets.

Time spent on the task



Time in seconds. Error bars are 95% c.i. for the mean

Experiment 2 : summary of results

Preferences, health, labels

- Simply asking subjects to choose healthy products results in an improvement wrt preferences
- Labels add to this shift considerably, especially for fat and vitamins

Diet tasks

1D Same performance, GDATL subjects slightly better

4D GDA = TL ; GDATL best & fastest at the same time.

7D GDA = TL ; GDATL best. Worse overall.

Thanks!

Data sources

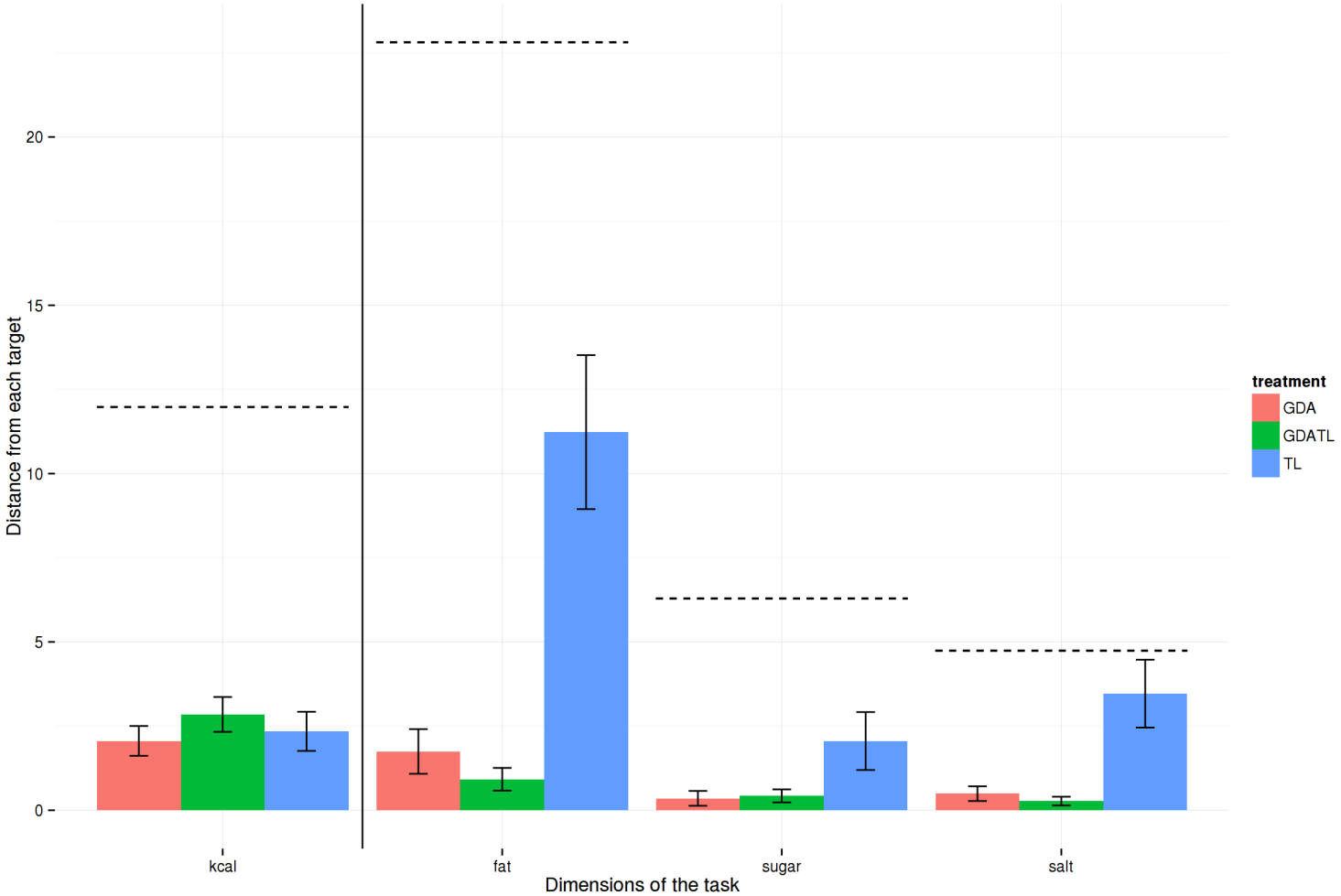
Product database

- 346 widely consumed products – full nutritional data
- SUVIMAX database (large household study in France in the 1990s)
- Missing data were inputted using :
  - Manuals and guides from SUVIMAX ;
  - SAIN/LIM (Nicole Darmon) ;
  - informationsnutritionnelles.fr

Computing TL and GDA

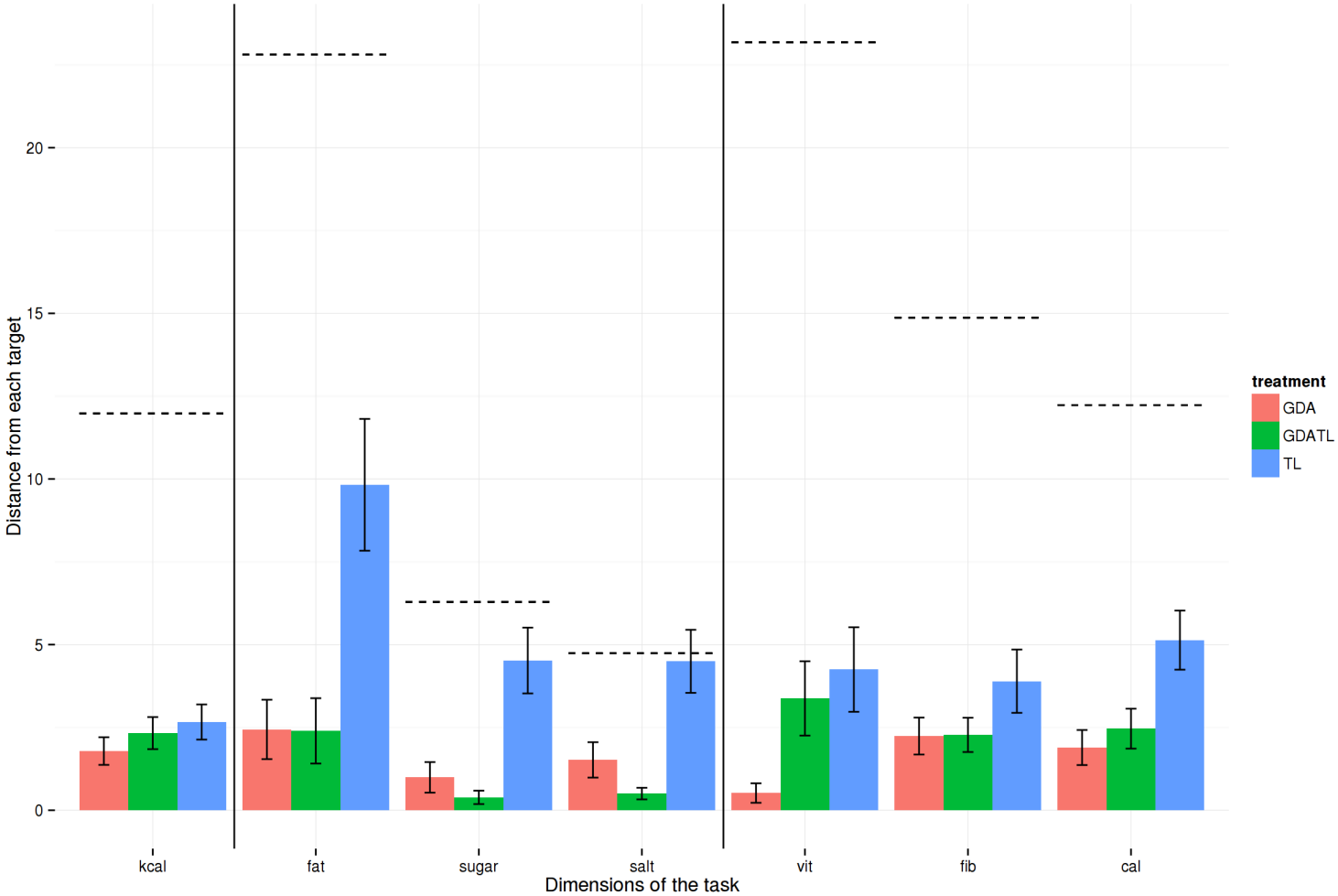
- GDA : EU Official Bulletin 22/11/2011 L304/61 Annex XIII, n° 1169
- TL : Simplified version of FSA TL definition

Breakdown of distance by target, 4D



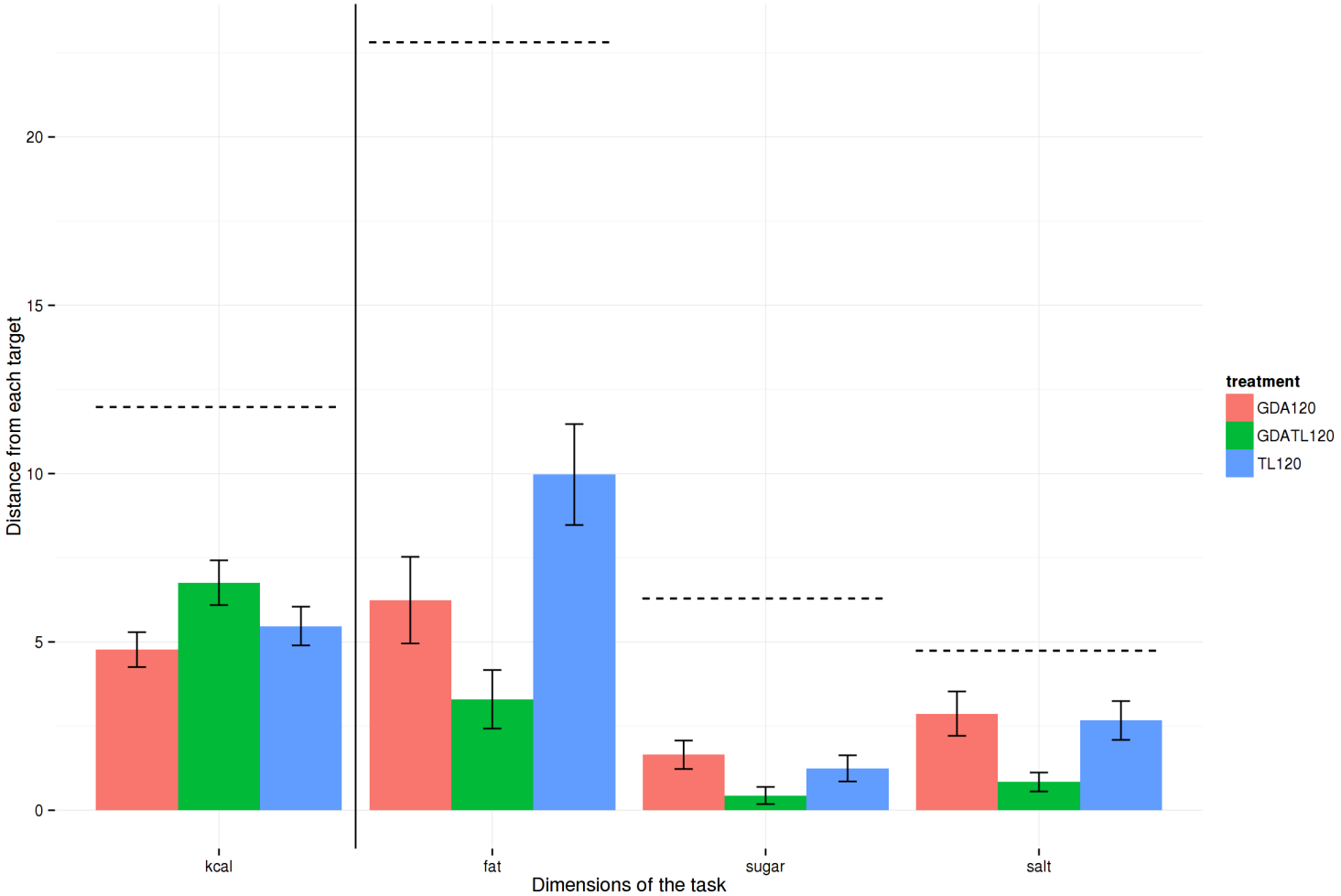
Dotted line is random play.

Breakdown of distance by target, 7D



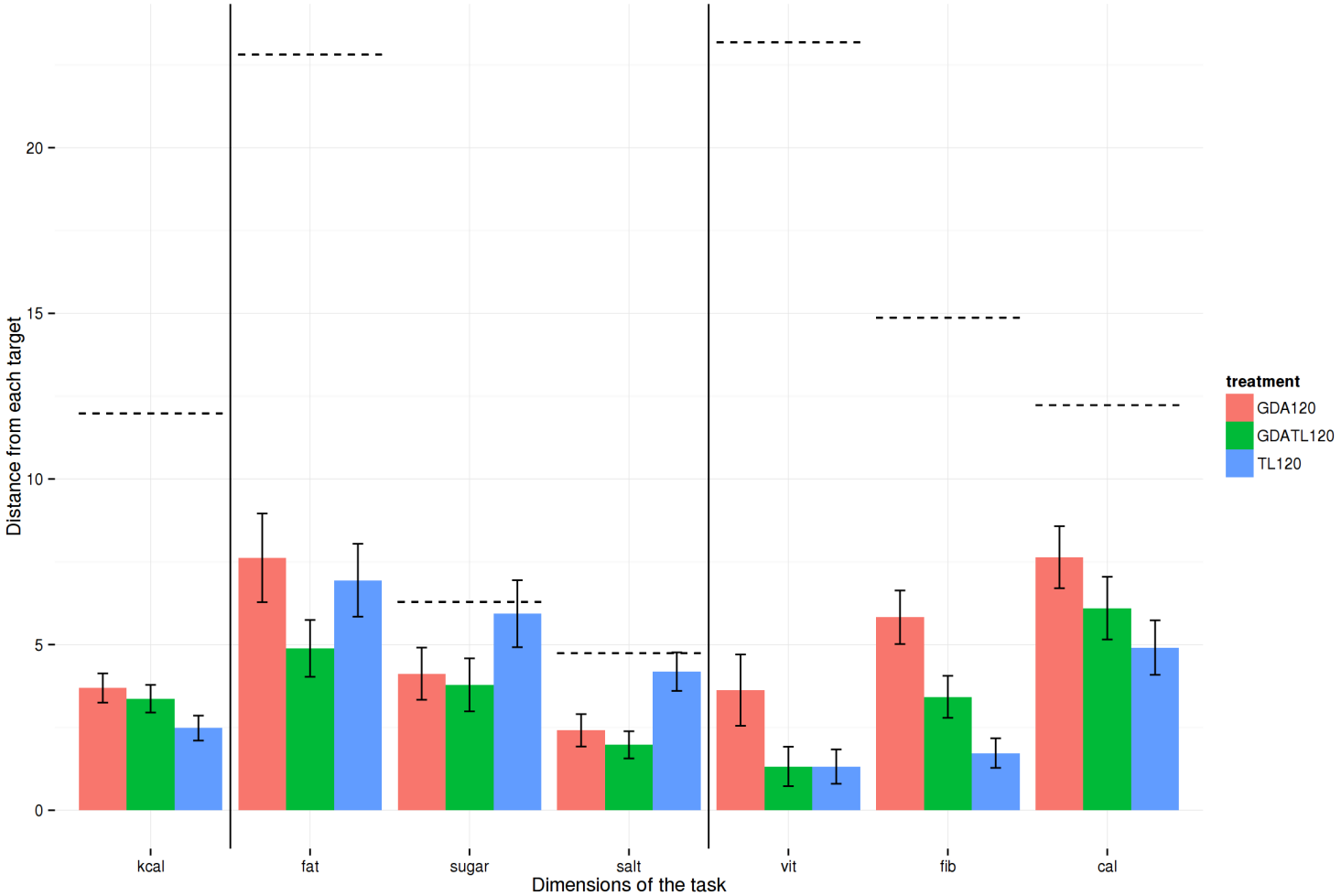
Dotted line is random play.

Distance by nutrient, 4D



Euclidean distance from target. 0 = on target on all targets.

Distance by nutrient, 7D



Euclidean distance from target. 0 = on target on all targets.

Regression analysis

	(1) correct		(2) distance		(3) time	
tl	0.0248	(0.74)	-1.478	(-1.29)	-3.307	(-1.16)
gdatl	0.0833*	(2.51)	-3.737***	(-3.31)	-6.992*	(-2.49)
d4	-0.339***	(-15.61)	9.347***	(13.64)	0.552	(0.37)
d7	-0.504***	(-21.18)	18.53***	(24.46)	3.828*	(2.33)
female	-0.0771**	(-2.61)	0.853	(0.85)	-0.917	(-0.37)
age	-0.00258*	(-2.01)	0.0672	(1.54)	0.0572	(0.53)
yearedu	0.0273***	(4.11)	-0.866***	(-3.85)	0.678	(1.21)
income	-0.00000750	(-0.48)	-0.000610	(-1.14)	-0.000750	(-0.56)
bmi	0.00377	(1.35)	-0.0773	(-0.82)	-0.418	(-1.77)
foodbudget	0.0000705	(0.29)	-0.00933	(-1.13)	-0.0277	(-1.35)
snacking	0.00132	(0.04)	1.946	(1.64)	1.681	(0.57)
_cons	0.322	(1.88)	17.30**	(2.98)	109.3***	(7.54)
N	2028		1961		2028	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table : Probit regressions