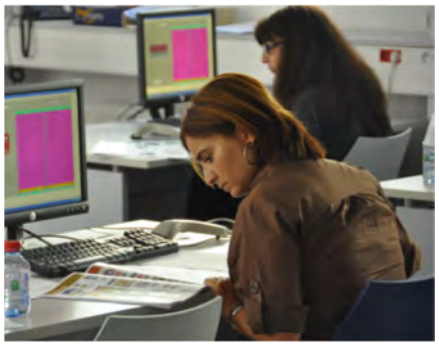
Observing purchasing behaviours

laboratory store vs. *field supermarkets*



Laurent MULLER

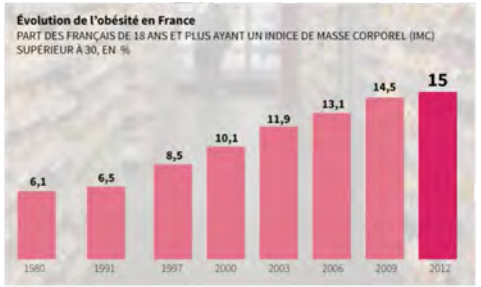
(GAEL INRAE)

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI RIMINI

Funded by the Joint Programming Initiative "A Healthy Diet for a Healthy Life" (JPI HDHL) with contributions from national funding agencies of participating countries

September 7th and 8th, 2021

Context: Obesity epidemic



Policy response

The Law of 26 January 2016 on the modernisation of the French health recommends a nutrition labelling system based on the nutritional composition of products.

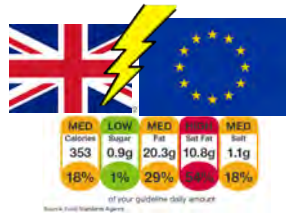
“The goal is to trigger a reflex: before buying, I look at the logo.”

« L'objectif est de déclencher un réflexe: avant d'acheter, je regarde le logo.»

French Health Minister Marisol Touraine

Context: A heated debate within EU

- In 2011, the UK Food Standards Agency unveiled the Traffic Light system. EU Commission states that it creates obstacles to trade, which violate EU law.
  The principle of mandatory labelling is abandoned in 2014.



« A comprehensive UK system of traffic lights food labelling is needed »

British Medical Association

Food industry

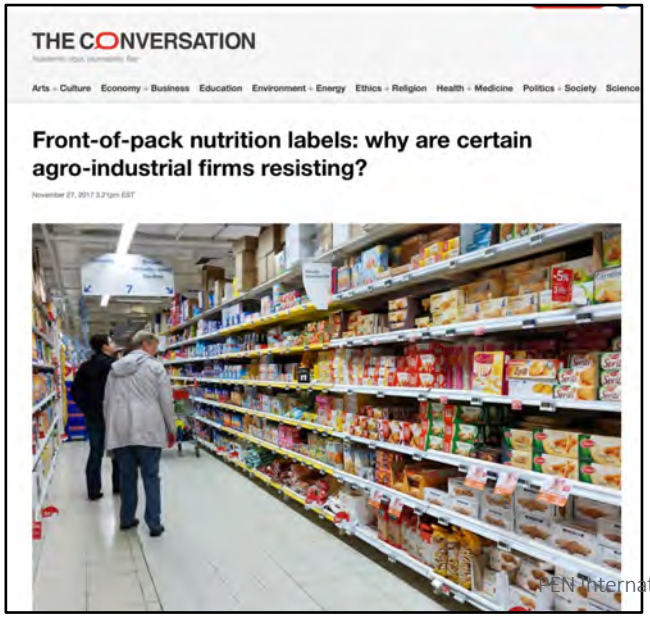
‘misleading’; ‘negative’; ‘overly simplistic’; ‘patronising’; ‘unscientific’

- In 2014, the French Ministry of Health proposes a simplified labelling system, the NutriScore, on a voluntary basis. EU response: “**Try it first**”.

Context: Also a controversial debate in France

M. Tourraine
Then Minister of Health

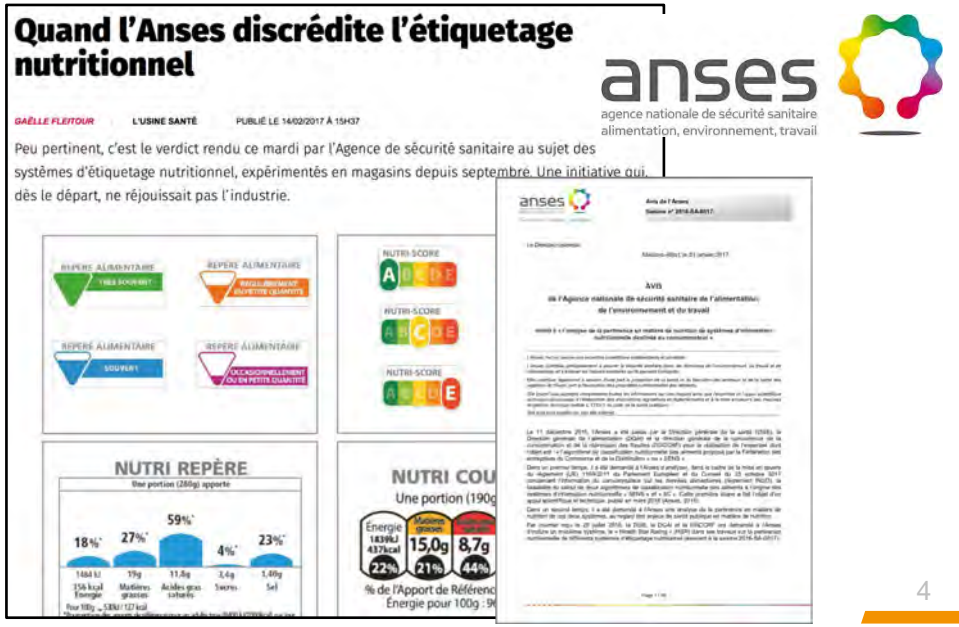
S. Le Foll
Then Minister of Agriculture


THE CONVERSATION

Arts · Culture Economy · Business Education Environment · Energy Ethics · Religion Health · Medicine Politics · Society Science

Front-of-pack nutrition labels: why are certain agro-industrial firms resisting?

November 27, 2017 3.21pm EST


Quand l'Anses discrédite l'étiquetage nutritionnel

GAËLLE FLEITOUR L'USINE SANTÉ PUBLIÉ LE 14/02/2017 À 15H37

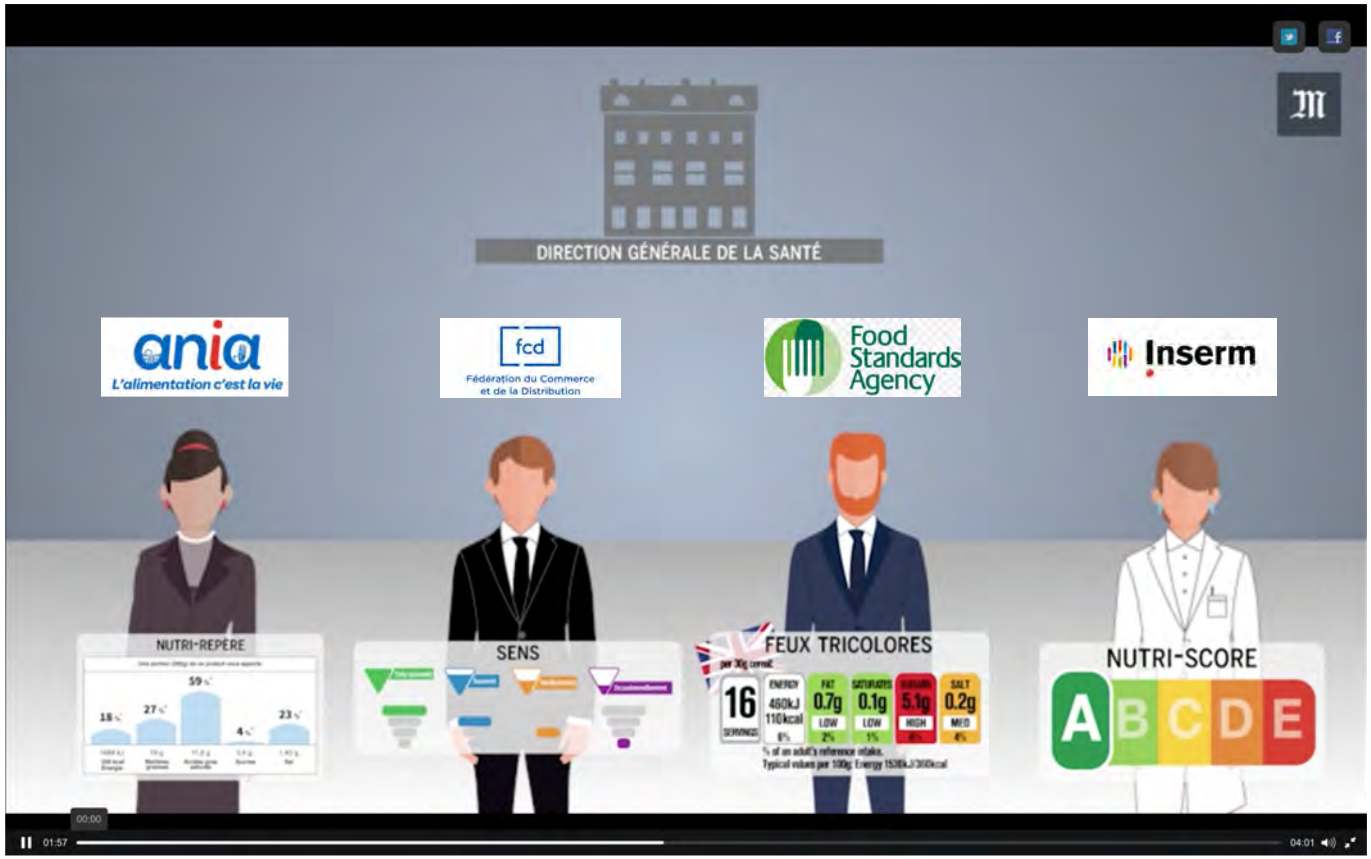
Peu pertinent, c'est le verdict rendu ce mardi par l'Agence de sécurité sanitaire au sujet des systèmes d'étiquetage nutritionnel, expérimentés en magasins depuis septembre. Une initiative qui, dès le départ, ne réjouissait pas l'industrie.

anses
agence nationale de sécurité sanitaire
alimentation, environnement, travail

NUTRI REPÈRE

Context: A trial to settle the question

France authorities decide to test **four** different nutrition labels during a trial period to see which one is the most efficient in encouraging consumers healthier food choices.



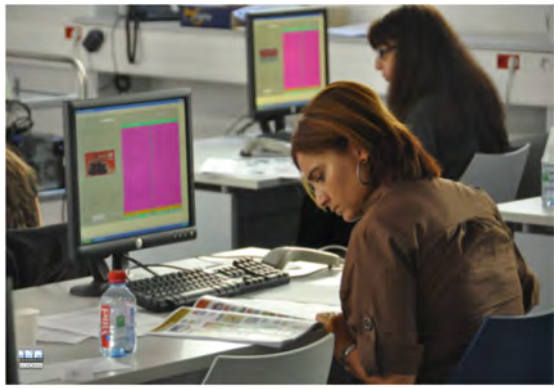
Context: Two studies are thus better than one

IN VIVO

Natural field experiment

BACK-UP PLAN

IN VITRO



Laboratory Framed field experiment

Context: Two studies are thus better than one

PEN Policy Evaluation Network

IN VIVO

ALiSS · Toulouse School of Economics · INSEAD

Journal of the Academy of Marketing Science
https://doi.org/10.1007/s11747-020-00723-5

ORIGINAL EMPIRICAL RESEARCH

Check for updates

Effects of front-of-pack labels on the nutritional quality of supermarket food purchases: evidence from a large-scale randomized controlled trial

Pierre Dubois[1] · Paulo Albuquerque[2] · Olivier Allais[3] · Céline Bonnet[1] · Patrice Bertail[4] · Pierre Combris[3] · Saadi Lahlou[5] · Natalie Rigal[4] · Bernard Ruffieux[6] · Pierre Chandon[2]

Received: 1 November 2019 / Accepted: 1 April 2020

Abstract
To examine whether four pre-selected front-of-pack nutrition labels improve food purchases in real-life grocery shopping settings, we put 1.9 million labels on 1266 food products in four categories in 60 supermarkets and analyzed the nutritional quality of 1,668,301 purchases using the FSA nutrient profiling score. Effect sizes were 17 times smaller on average than those found in comparable laboratory studies. The most effective nutrition label, Nutri-Score, increased the purchases of foods in the top third of their category nutrition-wise by 14%, but had no impact on the purchases of foods with medium, low, or unlabeled nutrition quality. Therefore, Nutri-Score only improved the nutritional quality of the basket of labeled foods purchased by 2.5% (−0.142 FSA points). Nutri-Score's performance improved with the variance (but not the mean) of the nutritional quality of the category. In-store surveys suggest that Nutri-Score's ability to attract attention and help shoppers rank products by nutritional quality may explain its performance.

Keywords Nutrition · Labelling · Supermarket · RCT · Food · Field experiment · Policy

Kelly Haws served as Area Editor for this article.

Pierre Dubois, Paulo Albuquerque, Olivier Allais, Céline Bonnet and Pierre Chandon contributed equally to this work.

The authors are grateful to the Editor, Associate Editor, and the Reviewers for their helpful comments and suggestions. They would like to thank Noël Renaudin, the Chairman of the scientific committee, as well as Benoît Vallet and Christian Babusiaux, the co-heads of the steering committee of the evaluation of graphical nutrition labels in real-life conditions launched by the French Minister of Social Affairs and Health. They also thank Daniel Nairaud, the managing director of the Fonds Français

Introduction

To promote healthier eating, regulatory authorities worldwide are encouraging the use of labels that provide simplified nutrition information on the front of the pack (FOP) in addition to the mandatory calorie and nutrition information already provided on the back. The European Union, for example, recently introduced a voluntary scheme for manufacturers to put graph-

IN VITRO

GAEL Grenoble Applied Economics Lab

European Review of Agricultural Economics Vol 47 (2) (2020) pp. 785–818
doi:10.1093/erae/jbz037
Advance Access Publication 21 August 2019

Nutritional and economic impact of five alternative front-of-pack nutritional labels: experimental evidence

Paolo Crosetto*, Anne Lacroix, Laurent Muller and Bernard Ruffieux
University Grenoble Alpes, INRA, CNRS, Grenoble INP, GAEL, Grenoble, France

Received May 2018; final version accepted May 2019

Review coordinated by Carl Johan Lagerkvist

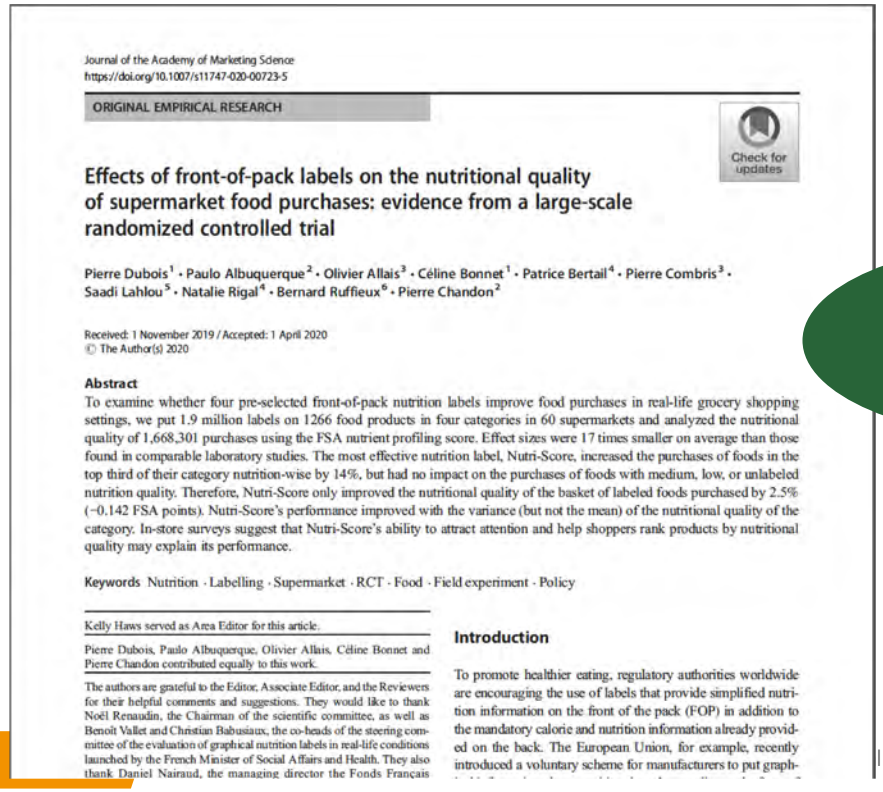
Abstract
An incentivised laboratory framed field experiment with 691 subjects examined the impact of five front-of-pack labels (Multiple Traffic Lights; Reference Intakes; HealthStarRating; NutriScore and Système d'Etiquetage Nutritionnel Simplifié) on food shopping within a catalogue of 290 products. Using difference-in-difference, we estimate the between-label variability of within-subject changes in the shopping's Food and Standards Agency aggregated nutritional score. All labels improve the nutritional quality (−1.56 FSA points on average). NutriScore is the most effective (−2.65), followed by HealthStarRating (−1.86). Behaviourally, subjects react mostly to the extreme values of the labels and not to intermediate values. Nutritional gains are not correlated with higher expenditure.

Context: Two studies are thus better than one

PEN Policy Evaluation Network

IN VIVO

ALiSS · TSE Toulouse School of Economics · INSEAD



Journal of the Academy of Marketing Science
https://doi.org/10.1007/s11747-020-00723-5

ORIGINAL EMPIRICAL RESEARCH

Check for updates

Effects of front-of-pack labels on the nutritional quality of supermarket food purchases: evidence from a large-scale randomized controlled trial

Pierre Dubois[1] · Paulo Albuquerque[2] · Olivier Allais[3] · Céline Bonnet[1] · Patrice Bertail[4] · Pierre Combris[3] · Saadi Lahlou[5] · Natalie Rigal[4] · Bernard Ruffieux[6] · Pierre Chandon[2]

Received: 1 November 2019 / Accepted: 1 April 2020

Abstract
To examine whether four pre-selected front-of-pack nutrition labels improve food purchases in real-life grocery shopping settings, we put 1.9 million labels on 1266 food products in four categories in 60 supermarkets and analyzed the nutritional quality of 1,668,301 purchases using the FSA nutrient profiling score. Effect sizes were 17 times smaller on average than those found in comparable laboratory studies. The most effective nutrition label, Nutri-Score, increased the purchases of foods in the top third of their category nutrition-wise by 14%, but had no impact on the purchases of foods with medium, low, or unlabeled nutrition quality. Therefore, Nutri-Score only improved the nutritional quality of the basket of labeled foods purchased by 2.5% (−0.142 FSA points). Nutri-Score's performance improved with the variance (but not the mean) of the nutritional quality of the category. In-store surveys suggest that Nutri-Score's ability to attract attention and help shoppers rank products by nutritional quality may explain its performance.

Keywords Nutrition · Labelling · Supermarket · RCT · Food · Field experiment · Policy

Kelly Haws served as Area Editor for this article.

Pierre Dubois, Paulo Albuquerque, Olivier Allais, Céline Bonnet and Pierre Chandon contributed equally to this work.

The authors are grateful to the Editor, Associate Editor, and the Reviewers for their helpful comments and suggestions. They would like to thank Noël Renaudin, the Chairman of the scientific committee, as well as Benoît Vallet and Christian Babusiaux, the co-heads of the steering committee of the evaluation of graphical nutrition labels in real-life conditions launched by the French Minister of Social Affairs and Health. They also thank Daniel Nairaud, the managing director the Fonds Français

Introduction

To promote healthier eating, regulatory authorities worldwide are encouraging the use of labels that provide simplified nutrition information on the front of the pack (FOP) in addition to the mandatory calorie and nutrition information already provided on the back. The European Union, for example, recently introduced a voluntary scheme for manufacturers to put graph-

Opportunity for methodological comparison

IN VITRO

GAEL Grenoble Applied Economics Lab

European Review of Agricultural Economics Vol 47 (2) (2020) pp. 785–818
doi:10.1093/erae/jbz037
Advance Access Publication 21 August 2019

Nutritional and economic impact of five alternative front-of-pack nutritional labels: experimental evidence

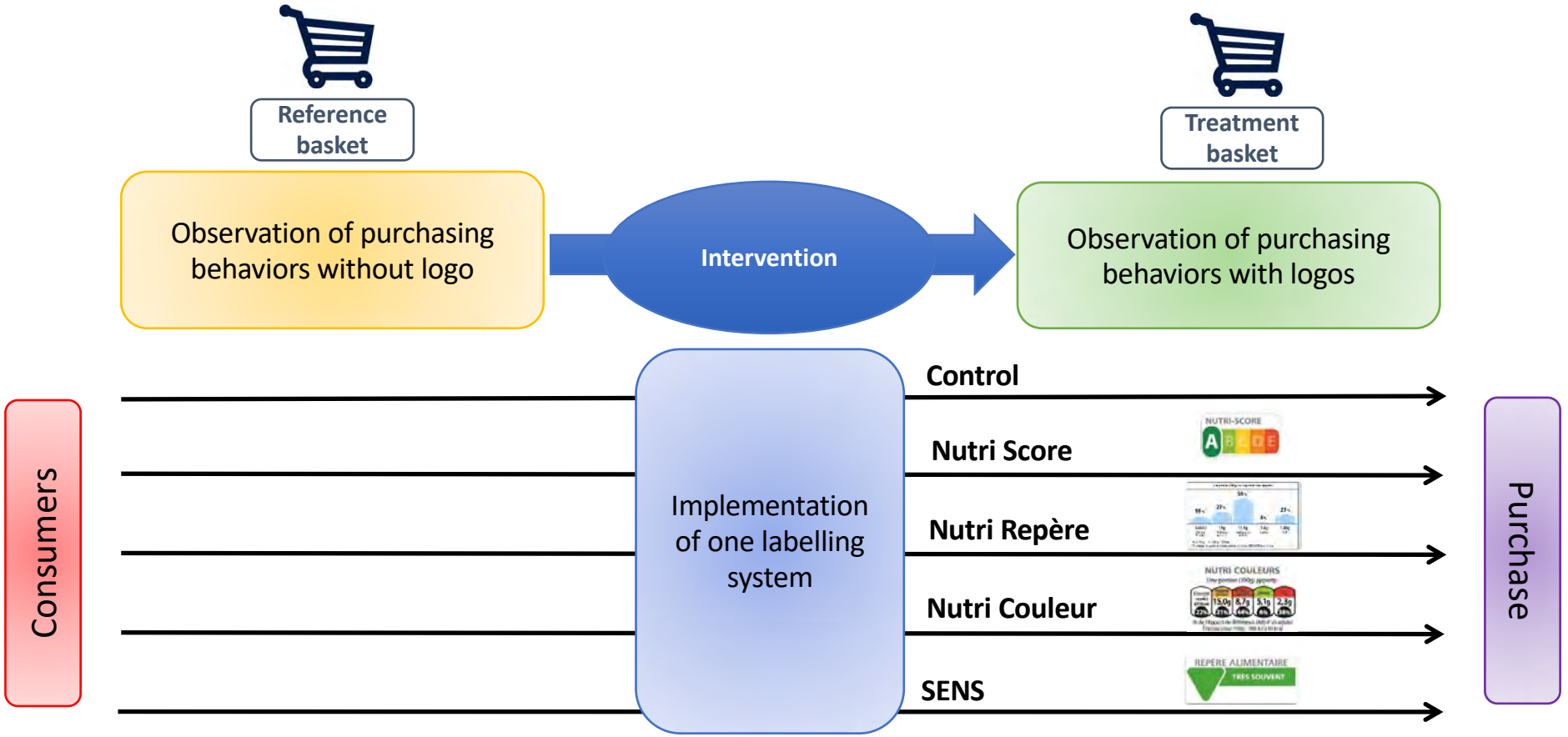
Paolo Crosetto*, Anne Lacroix, Laurent Muller and Bernard Ruffieux
University Grenoble Alpes, INRA, CNRS, Grenoble INP, GAEL, Grenoble, France

Received May 2018; final version accepted May 2019

Review coordinated by Carl Johan Lagerkvist

Abstract
An incentivised laboratory framed field experiment with 691 subjects examined the impact of five front-of-pack labels (Multiple Traffic Lights; Reference Intakes; HealthStarRating; NutriScore and Système d'Etiquetage Nutritionnel Simplifié) on food shopping within a catalogue of 290 products. Using difference-in-difference, we estimate the between-label variability of within-subject changes in the shopping's Food and Standards Agency aggregated nutritional score. All labels improve the nutritional quality (−1.56 FSA points on average). NutriScore is the most effective (−2.65), followed by HealthStarRating (−1.86). Behaviourally, subjects react mostly to the extreme values of the labels and not to intermediate values. Nutritional gains are not correlated with higher expenditure.

Method: Two identical experimental designs



Method: Two different modes of operation

Food STORES

IN VIVO

60 supermarkets

10 stores per systems + 20 stores for control

3 retailer brands

4 regions

10 weeks from September 26th to December 4th, 2016

IN VITRO

Experimental Platform in Grenoble INP

2 weeks from November 21st to December 2nd, 2016

51 sessions (1h30)

Method: Two different modes of operation

IN VIVO

Consumers' TASKS

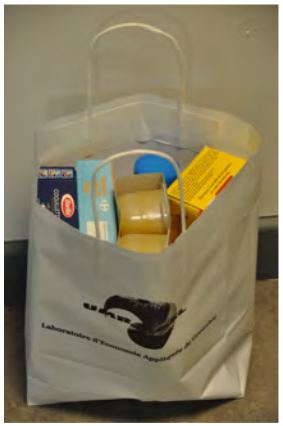
IN VITRO

No instruction

"Shop for your household for two days"

But:

At the end of the session, participant really purchase 1/4 of their food basket



Between phases, the labelling system was explicated using the same words

Method: Two different modes of operation

PEN
Policy Evaluation Network

IN VIVO

DECISION ENVIRONMENT

IN VITRO

Method: Two different modes of operation

PEN Policy Evaluation Network

PRODUCTS

IN VIVO

1266 products

Almost 2 millions logos (stickers)

Four shelves

Fresh prepared food – Canned prepared food – Pastries – Industrial breads

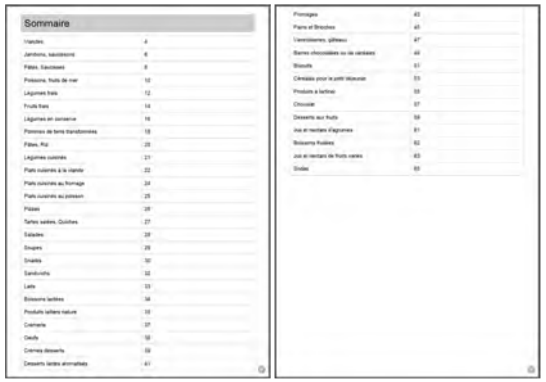
Logo coverage between 45% and 75%,

mostly retailer brand products.

IN VITRO

290 products

Across all food categories



All 290 products are tagged !*

* Except for ‘limited shelf with Nutri-Score’ treatment

Method: Two different modes of operation

IN VIVO

RESOURCES

IN VITRO

Coordination

1 Consulting firm

Labour force

60 labelling peoples

24 dietitians

Monitoring

6 auditors

Sciences

4 researchers for statistical analysis

6 researchers in the steering committee

4 researchers

1 software engineer

1 study engineer

2 assistants

Method: Two different modes of operation

IN VIVO	DATA	*IN VITRO*
171 827 consumers (loyalty cardholders)		**832 consumers**
1 668 301 purchases of 3586 products (of which 1266 were labelled)		27 882 purchases of 290 products (all labelled)
Outcome measures		**Outcome measures**
	FSA score normalised by 100 kcal	
Analysis		**Analysis**
	Difference-in-difference approach	

Results

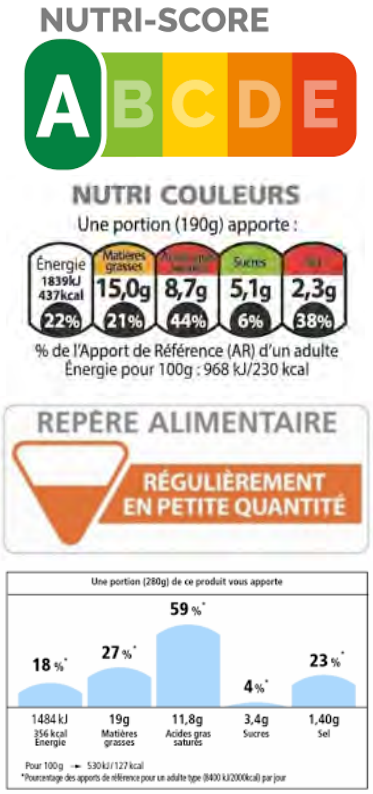
IN VIVO

1st. Nutri-Score

2nd. Nutri-Couleurs

3rd. SENS

4th. Nutri-Repère



IN VITRO

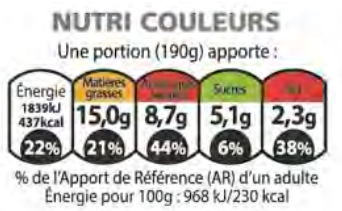
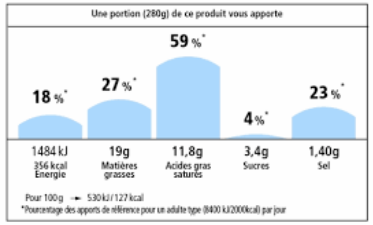
1st. Nutri-Score

2nd. Nutri-Couleurs

3rd. SENS

4th. Nutri-Repère

Results

	IN VIVO	FSA variation	IN VITRO
1st. Nutri-Score	-0.142*		-2.766***
2nd. Nutri-Couleurs	-0.115		-1.513*
3rd. SENS	-0.062		-1.140
4th. Nutri-Repère	-0.024		-0.924

Results

IN VIVO

1st. Nutri-Score

2nd. Nutri-Couleurs

3rd. SENS

4th. Nutri-Repère

FSA variation

-0.142*

-2.766***

-0.11

1.513*

-0.06

1.140

-0.024

-0.924

Correlation : 0.82

BUT

17 to 1 difference !

IN VITRO

1st. Nutri-Score

2nd. Nutri-Couleurs

3rd. SENS

4th. Nutri-Repère

Interpretation of the results

Why such a higher effect size in the lab?

The usual suspects:

HAWTHORNE EFFECT

STAKES

SAMPLE

CONTEXT

TASK

Interpretation of the results

IN VIVO **HAWTHORNE EFFECT** ***IN VITRO***

The nature and extent of scrutiny potentially influences behaviour

Consumers were informed of the local intervention through leaflets and totems

Consumers know all their decisions are registered anonymously

- Strategic behaviours ?
(Respondents send a message)

- Social desirability bias ?
(Respondents choose the most acceptable option)

Interpretation of the results

IN VIVO **HAWTHORNE EFFECT** ***IN VITRO***

The nature and extent of scrutiny potentially influences behaviour

Consumers were informed of the local intervention through leaflets and totems

Consumers know all their decisions are registered anonymously

- Strategic behaviours ?

If it exists, should be the same in both studies

- Social desirability bias ?

If it exists, should be the same in both studies

« *It is usually a minor problem in many experiments, especially if the decision envionment is interactive and 'rich', …, such as in market experiments* »

Falk and Heckman (*Science,* 2009)

Interpretation of the results

IN VIVO **STAKES** ***IN VITRO***

The stakes associated with laboratory experiments might not mirror those in play in the field.

Consumers pay what they buy **Consumers buy one quarter**

- Unfamiliar preference elicitation methods?
(e.g. Vickrey, BDM, etc.)
- Incentive compatible?
(make individuals act to their true preferences)
- Small stakes?
(motivate individuals to take the task seriously)

Interpretation of the results

IN VIVO **STAKES** ***IN VITRO***

The stakes associated with laboratory experiments might not mirror those in play in the field.

Consumers pay what they buy **Consumers buy one quarter**

- Unfamiliar preference elicitation methods?
  Not the case here
- Incentive compatible?
  It is costly to lie… not enough?
- Small stakes?
  True in both studies

« … *many decisions people make on a daily basis do not involve large stakes, implying that behavior in small-stakes experiments may be generalizable to these situations.* »
Charness and Fehr (*Science,* 2015)

Interpretation of the results

IN VIVO | **SAMPLE** | ***IN VITRO***

There might be differences in the type of people who participate in the lab and in the field

- **All cardholders from the treated supermarkets in 4 regions**
- **Underprivileged geographical area**
- **RCT on supermarkets**

- **Regular supermarket customers from the Grenoble metropolitan area**
- **Stratification by income**
- **Standard RCT**

- Differences in the subject pool?

- Self-selection bias?

(participants choose whether or not to participate; could lead to randomization bias)

Interpretation of the results

PEN Policy Evaluation Network

IN VIVO	SAMPLE	*IN VITRO*

There might be differences in the type of people who participate in the lab and in the field

- **All cardholders from the treated supermarkets in 4 regions**
- **Underprivileged geographical area**
- **RCT on supermarkets**

- **Regular supermarket customers from the Grenoble metropolitan area**
- **Stratification by income**
- **Standard RCT**

- Differences in the subject pool?

Participants in both study were similar in key characteristics (income, age, etc.)

- Self-selection bias?

Participants in the lab did not know the topic of the experiment (except that it concerns food purchase)… but still, they volounteer!!

Interpretation of the results

IN VIVO **CONTEXT** ***IN VITRO***

The context of the lab experiment differs from the field in ways that may influence behaviors

Usual shelves with stickers

Use of catalogues with prominent logos

- Is consumers' attention the same?

Interpretation of the results

IN VIVO **CONTEXT** ***IN VITRO***

The context of the lab experiment differs from the field in ways that may influence behaviors

Usual shelves with stickers **Use of catalogues with prominent logos**

- Is consumers' attention the same?

Definitely not!

« … *fine details of the decision context matter, such as the framing of a task or other factors that focus subjects on particular aspects of the problem. But carefully conducted laboratory studies offer far better controls of contextual factors relative to the field.* »

Charness and Fehr (*Science,* 2015)

Interpretation of the results

IN VIVO | **TASK** | ***IN VITRO***

To be comparable, the nature of the decision task in the lab must mimic the decisions made in the field.

- **Multiple purchase decisions over 5 weeks**
- **4 shelves, 3586 products, 1266 labelled**

- **Two consecutive purchase decisions**
- **All shelves, 290 products, all labelled**

- Time contraction

- Different scopes of measure

Interpretation of the results

IN VIVO | **TASK** | ***IN VITRO***

To be comparable, the nature of the decision task in the lab must mimic the decisions made in the field.

- **Multiple purchase decisions over 5 weeks**
- **4 shelves, 3586 products, 1266 labelled**

- **Two consecutive purchase decisions**
- **All shelves, 290 products, all labelled**

- Time contraction

Occasional needs, preferences may change
Low memory effect

Same utility function before and after
Encourage changes/substitutions

⇒ **Control generates saliency**

- Different scopes of measure

Interpretation of the results

IN VIVO | **TASK** | ***IN VITRO***

To be comparable, the nature of the decision task in the lab must mimic the decisions made in the field.

- **Multiple purchase decisions over 5 weeks**
- **4 shelves, 3586 products, 1266 labelled**

- **Two consecutive purchase decisions**
- **All shelves, 290 products, all labelled**

- Time contraction

Occasional needs, preferences may change
Low memory effect

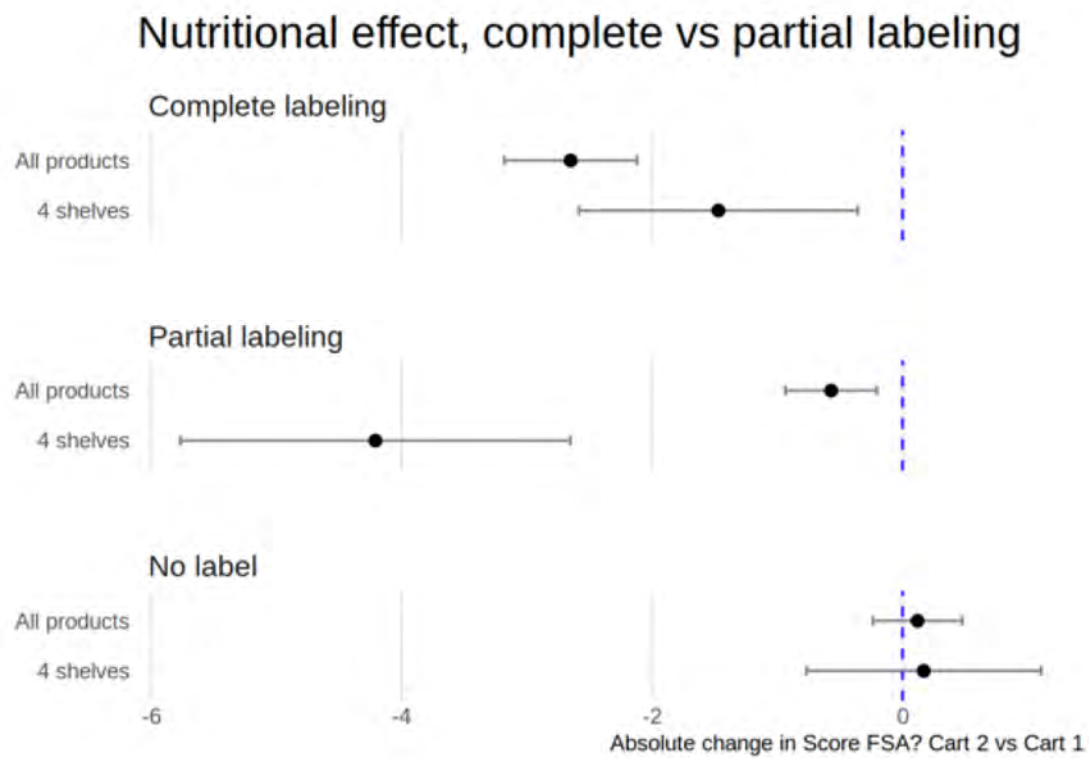
Same utility function before and after
Encourage changes/substitutions

⇒ **Control generates saliency**

- Different scopes of measure

Does it affect the results? Let's see…

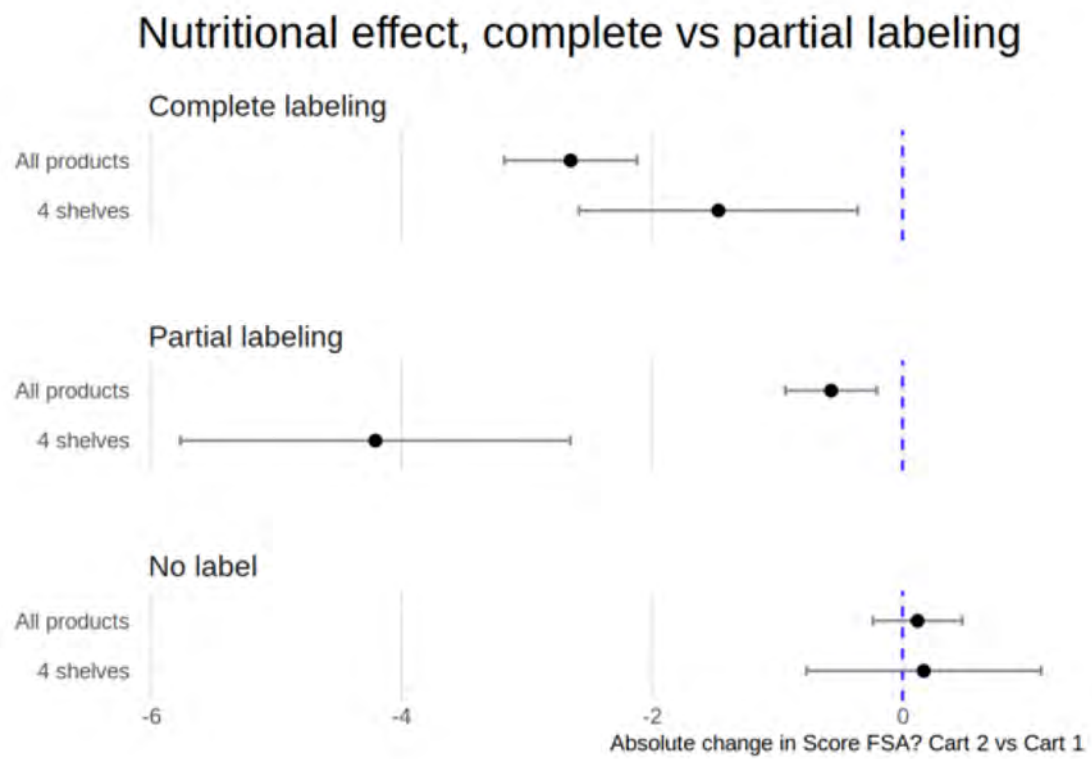
We actually test in the lab the impact of Nutri-Score on the same four shelves:



	All products	4 shelves
Intercept	5.109 ***	7.150 ***
	(0.651)	(1.972)
Basket 2	0.115	0.603
	(0.412)	(1.253)
Complete labeling	2.280 *	4.324
	(0.935)	(2.931)
Partial labeling	0.403	8.165 **
	(0.937)	(2.934)
Complete labeling X Basket 2	-2.766 ***	-4.229 *
	(0.591)	(1.866)
Partial labeling X Basket 2	-0.690	-6.892 ***
	(0.593)	(1.857)
N	702	372
R2	0.105	0.088
logLik	-1813.234	-1269.889
AIC	3640.468	2553.778

*** p < 0.001; ** p < 0.01; * p < 0.05.

We actually test in the lab the impact of Nutri-Score on the same four shelves:



	All products	4 shelves
Intercept	5.109 ***	7.150 ***
	(0.651)	(1.972)
Basket 2	0.115	0.603
	(0.412)	(1.253)
Complete labeling	2.280 *	4.324
Partial labeling		
Complete labeling X Basket 2	-2.766 ***	-4.229 *
	(0.591)	(1.866)
Partial labeling X Basket 2	-0.690	-6.892 ***
	(0.593)	(1.857)
N	702	372
R2	0.105	0.088
logLik	-1813.234	-1269.889
AIC	3640.468	2553.778

*** p < 0.001; ** p < 0.01; * p < 0.05.

Now,
48 to 1 difference!!

Interpretation of the results

IN VIVO | **TASK** | ***IN VITRO***

To be comparable, the nature of the decision task in the lab must mimic the decisions made in the field.

- **Multiple purchase decisions over 5 weeks**
- **4 shelves, 3586 products, 1266 labelled**

- **Two consecutive purchase decisions**
- **All shelves, 290 products, all labelled**

- Time contraction

Control generates saliency

- Different scopes of measure

The extent of effect size may be even worse… Another proof that attention is a key factor here!!

Interpretation of the results

IN VIVO | **TASK** | ***IN VITRO***

To be comparable, the nature of the decision task in the lab must mimic the decisions made in the field.

- **Multiple purchase decisions over 5 weeks**
- **4 shelves, 3586 products, 1266 labelled**

- **Two consecutive purchase decisions**
- **All shelves, 290 products, all labelled**

- Time contraction

Control generates saliency

- Different scopes of measure

The extent of effect size may be even worse… Another proof that attention is a key factor here!!

> « *What passes for "control" in laboratory experiments might in fact be precisely the opposite if it is artificial to the subject or context of the task.* »
>
> Harrison and List (*Journal of Economic Literature,* 2004)

Discussion

« *We argue that behaviour is crucially linked to not only the preferences of people, but also the properties of the situation.* »

Levitt and List (*Journal of Economic Literature,* 2004)

So why bother with laboratory experiments when it is impossible to perfectly reproduce the real world context?

Discussion

« *We argue that behaviour is crucially linked to not only the preferences of people, but also the properties of the situation.* »

Levitt and List (*Journal of Economic Literature,* 2004)

So why bother with laboratory experiments when it is impossible to perfectly reproduce the real world context?

Current consensus**:** Lab and field studies are complementary when lab experiments are conducted in order to tease apart potential confounds.

Instead, should we keep the laboratory to examine what cannot be examined in the field?

Discussion

« *We argue that behaviour is crucially linked to not only the preferences of people, but also the properties of the situation.* »

Levitt and List (*Journal of Economic Literature,* 2004)

So why bother with laboratory experiments when it is impossible to perfectly reproduce the real world context?

Current consensus: Lab and field studies are complementary when lab experiments are conducted in order to tease apart potential confounds.

Instead, should we keep the laboratory to examine what cannot be examined in the field?

But what about lab experiments aimed at evaluating the impact of policies?

Take-home messages

- **Control in lab comes at a cost**
  - Lab results better discriminate the impact of the competing labelling systems
  - Due to the increased consumer attention, the laboratory clearly overestimates the impact of the intervention.
  - *Quid* about possible underestimation of the field results (cofounding factors, poor attention, etc.)

    $\Rightarrow$To be verified with market data

- **Can lab capture quantitative effects? Apparently not. Does it matter?**

  (Contradicts Herbst and Mas (*Science*, 2015) who found no quantitative difference)

  – **No if the aim is to pick the 'best' option.**

  Lab studies act as magnifying glasses that are useful in distinguishing what intervention works better than another one.

  – **Yes if the aim is a cost-benefit analysis**

  Effect size is important when simulated results are used to assess future implications for society (*e.g.* epidemiology).

And even so… the wind tunnel of Schram (2005)

JPI

a healthy diet
for a healthy life

Thank you!

Determinants of diet and physical activity

Diet and food production

Diet-related chronic diseases