



HAL
open science

Experimental measures of behavioral effects: seven years of research on the road ahead

Paolo Crosetto

► **To cite this version:**

Paolo Crosetto. Experimental measures of behavioral effects: seven years of research on the road ahead. Economies et finances. COMUE Université Grenoble Alpes; Ecole Doctorale de Sciences Economiques, 2017. tel-02163273

HAL Id: tel-02163273

<https://hal.science/tel-02163273>

Submitted on 24 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE
UNIVERSITÉ GRENOBLE-ALPES
ECOLE DOCTORALE SCIENCES ECONOMIQUES

**Experimental measures of behavioral effects:
seven years of research and the road ahead**

Author:
Paolo CROSETTO

Supervisor:
Alexis GARAPIN

Jury:
Werner Güth
Guillaume Hollard
Stéphan Marette
Luís Santos-Pinto
Marc Willinger

*Document in support of the application for the
Habilitation à Diriger des Recherches*

in the

Grenoble Applied Economics Laboratory

November 20, 2017

Acknowledgements

There is a long list of people that made not only this work, but all of the works listed in this essay possible. It won't be possible to list them all, but I'll try.

Thanks to Alexis Garapin, who accepted to be my supervisor for this mystery object that was the HDR to me. In the process, Alexis had to bear with my delays and last-minute-panic attitude, was patient and persistent in reminding me what lied ahead, and supportive throughout. Thanks, I owe you one.

I wouldn't be a researcher if several people had not believed in me and urged me to continue along the academic path. This involved believing in me even at times when I was losing faith myself. So I cannot but owe the deepest gratitude to Alessandra Venturini, that nudged me towards research right after my undergraduate studies, and to Stefano Sacchi and Maurizio Ferrera, with whom I worked at URGE and convinced me to go for a PhD. To John Hey, that infused me with hope and enthusiasm during my PhD, and allowed me to join his team in LUISS. To Werner Güth and all the colleagues at MPI Jena, that gave me the life-changing opportunity to work in one of the best Experimental Economics groups in Europe. And finally to the colleagues at INRA, who hired me and gave me full support and complete academic freedom, together with exciting new areas to explore.

Research projects are joint efforts. This is particularly true with my coauthors, since we tend to disregard centuries of economic thought and *do not* practice the division of labour – everyone does (and re-does, and re-does) everything. So the works I'll be talking about here exist because of these somewhat irrational joint ventures with several brilliant people around Europe. Thanks to, in no particular order, Antonio Filippin, Alexia Gaudeul, Gerhard Riener, Tobias Regner, Luigi Mittone, Matteo Ploner, Werner Güth, Maria Vittoria Levati, Julia Brüggemann, Lukas Meub, Kilian Bizer, Bernard Ruffieux, Laurent Muller, Anne Lacroix, Ori Weisel, Fabian Winter, John Hey, John Bone, Marco Mantovani.

Finally, I have to thank my family. This is weird, because family care, and especially having to deal with the hundred ways in which Pietro manages to get hurt and cause havoc each and every day – took so much time and energies that if anything it severely delayed my work. But nonetheless, or for this very reason, thank you Pietro, thank you Irene, for infusing me with joy and love and tenderness and being there for me at any time. And thanks to my parents, for having brought me up the way they did, giving me that most precious gift of knowing you'll never walk alone. I hope to be able to give to Pietro that very same gift.

Contents

Acknowledgements	i
Why an HDR?	1
What did I do since my PhD?	2
Recurrent themes	2
Public goods and the Internet	4
Risk preferences	13
Context effects	29
Food choices	36
Where do we go from here?	44
Cognitive limits, context effects and the challenge of convergence	44
Cautious incursions in the land of neuroeconomists	46
Introducing social pressure in food WTP	48
Eliciting fast & slow responses	49
Farther along the way	55
What could I teach to students?	57
Bibliography	60
Appendices	71
Appendix A Curriculum Vitæ	72
Appendix B Publications	77

Why an HDR?

Not being French, until recently I did not know what an HDR – Habilitation à Diriger des Recherches, accreditation to supervise research – even was. I have a clearer idea now: it is the degree that entitles a researcher to officially supervise PhD students. Armed with this knowledge I can try to introduce this work with the most important question of all: why would I want to supervise research?

The most obvious reply is that this is the natural next step in a researcher's career. Like in most other trades, first you learn it, then you practice it for a while, and then you teach it to newcomers. The second, practical goal in having students is that as the research projects grow in ambition and scope, more than one person needs to be involved, and young PhDs are the backbone of most research programs. But for me, these sensible and obvious reasons were not enough – after all, a supervisor wields a lot of power and by his choices can influence the entire careers of his students. One has to be really willing to do that.

There are two more important and personal reasons for my HDR. The first, is that what I like most about our job is the possibility to exchange ideas with clever people and keep learning from others. In the few opportunities I had to interact with students approaching research, I learned a lot. I'm eager to learn more. The second, is that I think I could really help young PhD students in finding their way around. Not because I now know a bit more about research than when I finished my PhD, but mainly because during my PhD I was frequently lost, and I know what it feels like – the memory is still fresh.

In the remainder of this essay I will try to convince you – my HDR committee – that I am indeed able to supervise PhD students and to help them in shaping their research interests into an academic career. This will be done in three steps.

First, I'll summarise what I have done so far. I have tried to be brief, but research results require context, and context requires details. I tried to give a bird's eye view, packing together similar papers, and giving some perspective about my overarching research goals rather than just piling up papers.

Then, I'll briefly expose the main lines of my ongoing and future research, focusing on what keeps the whole story together. In doing so, I'll try to show which research projects more than others might see the inclusion of a PhD student, since they could be both good avenues *into* the research craft and *out towards* a worthy contribution to economic knowledge and a nice career.

Finally, I'll ask myself what could I teach to students, and if I'd make a good supervisor. This is the shortest part, since I frankly know little about that. But I tried to ask myself that question, and I think it worth the while to share my thoughts with you.

Sit tight: we start off.

What did I do since my PhD?

Recurrent themes

The report on my research activity and results you are about to read is of a rather differentiated nature. You will read about open source software, copyright, crowdfunding, partnerships, cooperation, punishment, risk elicitation, gender, context effects, industrial organization, and food choices and labeling. And this is just the list of topics I worked on since starting the PhD. Before that, I studied international politics, and then the Russian transition from plan to market. I had two work experiences as an early junior researcher: one on the economics of transition and the other on European governance and social welfare issues.

Most academics choose a topic and stick with it in the long run. Not me. I have always loved diving into new subjects, and this is what I do in my research activity. If I haven't jumped discipline boundaries since my PhD, it is just because I am not good enough to do it. I change topic, I change my mind, I make a contribution and move on. This is not terribly efficient, but it is the only way I can satisfy my curiosity and be (somewhat) productive.

Yet, I think there are some common threads underlying my research activity. I list them here with the aim of guiding you through the at first sight incoherent set of results that lies ahead.

The first and most obvious common thread is the *experimental method*. I got to learn about the existence of experiments in economics during my first year of PhD, thanks to a course made by Miguel Costa-Gomes in Milan. Experiments gave me what I was looking for: a look at behavior, a challenge to the somewhat aseptic topics of the Mas-Colell, Whinston and Greene *Advanced Microeconomics* book, a hands-on approach, and a feeling of science advancing, one experimental manipulation at a time. All but one of the papers discussed here are experimental. All of my near future projects are.

The second common thread is a focus on *individual decision making*. This might seem obvious to economists, who mostly are methodological individualists. But when starting my PhD I came from political science, I had studied lots of law and sociology, and most of the economics I had studied and practiced was macro. I looked at the world from a macro perspective. Social institutions for me just *existed* and I was concerned with understanding their mechanisms and interactions; the idea that institutions of all kind are the results of individual decisions, that they are *generated* – in complex ways – from these very decisions, had literally *never* occurred to me. It was an epiphany. Some papers in particular were game changers: Schelling, 1971 segregation model, Anderson, 1972 Science paper on *more is different*, Axelrod, 1984 Prisoner Dilemma contests and several papers from Evolutionary Game Theory. All my research works came out of this changed perspective, so they not only all share a microeconomic focus, but have a particular focus on individual heterogeneity. People are different. I see most of my research efforts

as a way to better identify and measure the extent of this individual heterogeneity, so that we can better map individual behavior onto social issues.

Consequently, another common underpinning of my works is *measure*. I strive in my research to *measure* effects, as opposed to just *showing* that a phenomenon exists. While some might argue that science itself is measurement, I have a feeling that in economics we have not been for a long time measuring enough. Economic courses focus on the concepts rather than applications; most concepts are unobservable; and theorists strive for generality and shy away from measure. To be sure econometrics is by its very name the attempt to measure economic parameters, so I would not claim to be alone in my measuring efforts. But I do think that most early experimental economics has been concerned with testing the limits of economic theory, and for this a limited sets of counter-examples does the job. I think that the next step for behavioral and experimental economics is to move from showing that behavioral biases exist to measuring them, in order to assess their practical relevance. This is at play in risk elicitation, of course, but I tried to measure effects – if possible at the individual level, to account for heterogeneity – in every field in which I worked.

The last but I think most important recurrent theme is methodological innovation. Partly because I get easily bored, partly because I tackled some questions outside of the path trodden by experimental economists, and also because of a willingness to measure or better measure things, I often engaged in creating *new methods*. For my PhD thesis I developed a variant on the Scrabble game to measure the use of different copyleft licenses, than later evolved into a tool to measure the impact of different intellectual property regulations. Later, with Antonio Filippin, I developed the Bomb Risk Elicitation Task. Roughly at the same time, with Alexia Gaudeul, we came up with a new way to measure context effects at the individual level. After joining the Grenoble group, we came out with several new experimental methods to test different aspects of food choices. In all projects I strove to develop a tool that was closely tailored to the research question, rather than relying on standard workhorses.

With these recurrent themes in mind, I will now expose the different domains in which I have been doing research during and since my PhD, in no particular order. For each theme, I will give a brief introduction to the topic, to its main points of interest, and to my contributions, including published or submitted papers only. All future research perspectives, as well as ongoing and unpublished projects are discussed in the next chapter.

Public goods and the Internet

My PhD thesis started out of a deep interest in Open Source software. It fascinated me how highly complex products like a web browser, an operating system, and several hundred other software projects could be the result of uncoordinated cooperation among strangers over the Internet. I did not frame it like this back then, but in a nutshell my question could have been put this way: *does the existence of the Internet facilitate or hamper cooperation in public good dilemmas?*

There are arguments for both answers. On the one hand, since the cost of communication and link creation has dramatically gone down – and so has marginal cost, that has dropped to zero for software or anything else that can be turned into a digital artifact – then cooperation should increase. It is easier to communicate, it is easier to meet, it is not costly to share. The rise of the Open Source software movement is a monument to this new surge in cooperation. At the same time, game theory teaches us that repeated interactions are key to the building of trust. And with an on-line world connecting hundreds of millions of individuals in real time, one-shot interactions become ever more common. Opportunities for free riding blossom, as rampant piracy in the 90s and 2000s has shown. Is the Internet facilitating or discouraging cooperation and trust?

This underlying question gave rise to three distinct set of papers. One focused on the role of Intellectual Property Rights in highly sequential innovations like the software industry. Another dealt with more classic questions in public good research, like punishment and cooperation, but with a focus on the drivers of (in)stability of partnerships in dyads (co-authors, co-developers in open source projects) and in the motives of sanctioning. A third, developed with Tobias Regner, deals with entrepreneurial finance, the rise of crowdfunding platforms, and the interaction between donors' motives and entrepreneurs' behavior. In the following, I'll give some details about these three streams of research.

Experiments on innovation and copyright in a sequential setting

If I have seen far, it is because I was standing on the shoulders of giants.

Isaac Newton

One area in which this key question is central is intellectual property and its evolution in the digital age. Intellectual property (IP) is a way to solve the dynamic inefficiency inherent in the production of intellectual work and ideas – essentially, public goods – by introducing a (temporary) static inefficiency: monopoly power. Granting temporary monopoly rights on ideas, innovations, books, or software provides incentives to creators by protecting them from imitation and allotting to them a part of the social surplus generated by subsequent innovators (Arrow, 1962). On the other hand, the creation of monopolies on innovations increases prices, distorting resource allocations, causing inefficiencies and leading to welfare losses (Boldrin and Levine, 2008b; Stiglitz, 2007).

The debate around IP started right after the first IP laws were passed, in the eighteen century. The advent of the Internet gave it a further spin. On the one hand, the availability of unlimited costless copying meant that doctrines like fair use or the right to sell a used copy of an object

could be extended in such a way to actually make the whole concept of IP meaningless. On the other hand, the creators, through the use of aggressive licensing or technological devices, could effectively lock up content in a much more closed and secure way that had ever been the case. This new set of arguments made it to the center of the stage in the late 90s and early 2000s, when a host of political, legal and economic works analyzed what came to be called a 'copyright war' (Benkler, 2006; Lessig, 2001).

In my PhD thesis I concentrated on a particular setting: sequential and cumulative innovations. Most of today's products are made of several parts, each of which is the result of the combination of different sets of expertise, methods, inputs, and machinery. Most innovations are indeed cumulative: they do not come out of the blue, but build on some other, previous innovation. Academic research is quintessentially sequential; and so are software, biotechnology, pharmaceuticals and information technology. The key issue with sequential innovation is that strong IP rights upstream can result in reduced opportunities and inefficiency downstream (Boldrin and Levine, 2008a; Heller and Eisenberg, 1998; Scotchmer, 1991). There will be less innovation if the upstream innovator appropriates all of the rights that make sequential innovations possible.

A theoretical paper by Bessen and Maskin, 2009 shows that, when innovations are sequential and complementary, then not only it is in society's best interest to abolish IP rights; but, in some cases, it is in the *firms'* best interest *not to* apply for IP protection of their innovations. The intuition is simple: by opening up an innovation, the firm makes further innovations possible. To the extent that those enlarge the market, everyone is better off. It is the rationale behind a counterintuitive advertisement campaign launched by Steve Jobs for Apple in 1981, running 'Welcome IBM, seriously', and saluting the debut of IBM's first personal computer – a firm spending money to praise a competitor. The presence of more PCs would enlarge the market, spur further innovations, and, in the end, benefit everyone.

This paper was enlightening for me, as it rationalized in economics terms the practice I saw around me everyday as a Linux and Open Source enthusiast: commercial companies giving away their code for free, and whole operating systems and applications (like the ones I am using right now to type these words) being born out of that. I hence decided to do the most naive thing possible – being a PhD student, that was allowed – and just 'test the findings in the lab'. I know now that testing theories is not so straightforward – theories are by definition true, and you can test the robustness and real-world validity of their axioms, or test theory robustness by relaxing some of them, or you can use them to set up benchmarks against which to observe behavior. But back then I just wanted to see if I could reproduce in the lab a pattern that I saw in real life – firms, including big players like IBM and even, at a later point, Microsoft, increasingly moving to open source strategies.

In deciding to bring the dynamic, creative world of innovation in the laboratory I faced a central issue: to increase internal validity, I would have to keep as close as possible to the theoretical model. But to increase external validity I had to infuse the experiment with at least part of the context of innovation. This is not trivial. Innovation happens because of a mix of skill, creativity, effort and investment; it happens over time and in a discontinuous fashion; moreover, innovators have skin in the game and have a feeling of ownership and entitlement over their creations that goes beyond those of a simple investment (Buccafusco and Sprigman, 2011).

How do you translate this in the lab? Some experiments chose to model the creative process using search over complex spaces. Subject explore a space looking for some optimal solution

that yields higher payoffs, and that the experimenter knows and controls. Often this optimal solution is randomly chosen by the experimenter over the space (Buchanan and Wilson, 2012; Cantner et al., 2009; Dimmig and Erlei, 2012; Ederer and Manso, 2013; Meloso, Copic, and Bossaerts, 2009). By choosing to implement search tasks, these experiments abstract away from the crucial features of creativity and individual skills. Finding the ‘right’ combination is often just a matter of luck and enough trials. A smaller set of papers choose instead to implement outright “creative” tasks, i.e. writing poems, or painting, tasks involving creative skills rather than the search (Bechtold, Buccafusco, and Springman, 2015; Buccafusco and Sprigman, 2010, 2011). Such designs capture the creativity core of innovations better, but forfeit control – it is impossible to objectively assess which poem is ‘better’ or ‘more creative’ in the set.

In Crosetto, 2010 I chose to use *Scrabble* to create a task integrating both creativity and control of the outcomes. In the task subjects innovate over a familiar space (their language), using both economic (experimental money) and cognitive (creative effort) resources. The production process requires creativity and skill, but its results are countable and can be used to create precise statistics and comparisons across conditions. By allowing subjects to extend already created words, the design induced sequentiality and dynamics in an intuitive way. Subjects face a trade off between creating short words (roots) that could be profitably licensed to subsequent innovators, or extend existing words, but paying the fees involved.

In this first *Scrabble* paper subjects had the strategic choice of submitting their work to IP rights, that is, asking for a license fee for any extensions; or submitting them to *copyleft*, that is, using an open source license granting free access but also not allowing any follow-up innovator to privatise the ensuing extensions. Treatment variables were the exogenous, fixed license fee (high or low) and the possibility to anagram words rather than just extend them adding one letter in any position (the default). The first manipulation gave insights about the effect of up-stream cost on downstream innovation rates; the second, the effect of an explosion of extension possibilities on license choice.

The key variable of interest was the emergence of *copyleft*, i.e., the number of subject choosing *not to* subject to IPR their works, and the productivity of each IP mode, *copyleft* and *copyright*. As seen in the real world and as predicted by Bessen and Maskin, 2009, open source use did emerge, and particularly so when royalty fees were relatively high – for subjects prefer to avoid paying for license fees and *get* the intuition that *copyleft* extends the market – and when the extendability, modularity and manipulability of inventions is enhanced – since the future expected stream of extensions is much larger, and so the benefit of a free sharing of roots are more easily observable.

That paper was published in the published proceedings of an IT conference – something usual in IT research – not in an economics journal. This hampered my efforts to publish it in economics – a naive mistake from which I learned a lot. But not all hope was lost. In a serendipitous twist of fate, a seminar I gave in Göttingen spurred the interest of a PhD student there, Julia Brüggemann, and we joined efforts to come up with a new *Scrabble* paper that would use the task developed to ask a more policy-relevant question: what is the impact of the existence of IP rights on innovation and welfare? The experimental tools I had developed were well suited to build a *counterfactual* situation, i.e. create in the lab a society without IP rights that could be in all other aspects identical to another with IP.

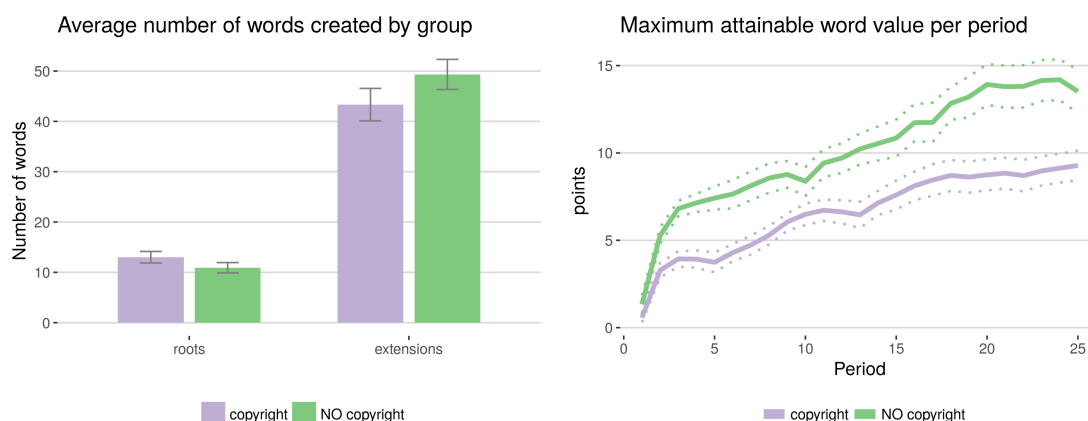


FIGURE 1: Drivers of detrimental effects of IP rights in Brüggemann et al., 2016

Instead of focusing on the endogenous choice of subjects to go for copyright or not, in a new paper (Brüggemann et al., 2016) we focused on the consequences of different institutional arrangements. IP rights were either allowed – and imposed by default on all new words, as it is the case in the real world – or not allowed – mimicking the state of affairs before the Statute of Anne first introduced copyright in 1710.

We found intellectual property to have a sizable and robust adverse effect on welfare. This is mainly due to two drivers. First, introducing intellectual property results in more basic innovations. Subjects have at any point in time the choice between creating a new three-letter word (named ‘root’) or extend an existing word. Extensions give in general higher profits, since, as in standard Scrabble, a player gets the value of all the letters of the word and not only of her marginal contribution. But when introducing copyright, extenders have to pay a license fee to the word owner. Introducing copyright shifts the most widespread strategy from extensions to the creation of roots in the hope of getting a revenue stream from license fees. Consequently, more basic, three-letter words and less high value, longer words are created when IP is introduced (Figure 1, left). Second, the struggle to avoid paying license fees results in foregone opportunities. Subjects fail to exploit the most valuable sequential innovation paths, so much so that the *best* word they can create by fully using their cognitive skills in each period under IP rights is, by the end of the experiment, worth the same as an *average* word created with little effort in no-IP treatments (Figure 1, right).

This experiment proved me that experimental economics could venture on new grounds and dare to ask policy-relevant *macro* question. Of course, the external validity of the finding might be doubted – it is just students playing with words. But the ability to create counterfactual situations and to investigate the *mechanisms* driving the results is unique to experiments. Coupled with a clever use of the field, laboratory experiments can shed light also over issues at first sight far removed from the core business of experimental economists.

Cooperation, punishment and partnerships in dyads

Exit, voice and loyalty

Albert O. Hirschman

A second front of research born out of my interest in open source software and group effort has dealt with the more fundamental question of public goods. I fear all experimental economists have sooner or later run one or more public good experiments; and I am no exception. The work done in this area had two main twists, both along classic Hirschman, 1970 lines of *exit* and *voice*.

Inspired by the dynamics of cooperation in joint ventures in the open source world and in academia, together with Gerhard Riener and Alexia Gaudeul we investigated the effect on partnership duration and efficiency of varying the availability and cost of outside options in Gaudeul, Crosetto, and Riener, 2017.

When you think about joint projects, for instance in co-authorship, you face several key problems. First, outcomes are to some extent stochastic. As every academic knows, publication is not a deterministic function of paper quality. Surely higher quality increases the chances of publication, but does not ensure publication. Second, most of the time you cannot strictly monitor the quality of your coauthor's work. We specialize in different things, and I cannot really say if the subtle econometric specification my coauthor came up with is really as good as she says, as well as she cannot really judge my experimental design decisions. In some cases, as in open source remote cooperation, you have even less of a clue of your peer's output quality. Third, you all work on several projects at any one time, and you do not know how high or low *this* very project ranks in your coauthor's priority list. Your coauthor can at any time move to some other projects and leave you stranded with an unfinished, unsubmitted, or freshly rejected paper. I know all this, since I have been on both ends of these stories, acting both as the good or the bad guy; you also do know, I'm sure. Similar stories could be told of other dyads, like marriages.

In the paper we set up a public good experiment with stochastic outcomes, imperfect monitoring and an outside option. We vary the attractiveness of the outside option and track its consequences on the likelihood of the dyad to explode and on the efficiency of surviving dyads. Crucial to our results, we elicit beliefs about the actions of the peer; these beliefs give us the ability to pin down the main motives of the break down of dyads.

There is some literature on the impact of outside option on partnerships and joint efforts. Some argue that it should be easy to exercise an exit option. Allowing exit may provide a way out from dysfunctional pairs (MacLeod, 1993), the threat of exit may be necessary to enforce cooperation (Lin, 1990); allowing exit may help to frame the collaboration as the product of the free will of both participants (Yamagishi, 1988), enhance trust (Bravo and Squazzoni, 2013) and thus promote intrinsic motivations and cooperation and promote efficient self-selection (Orbell, Schwartz-Shea, and Simmons, 1984). Opposite arguments exist: exit implies that a peer can easily escape punishment for low contributions (Dong and Dow, 1993; MacLeod, 1988), it might encourage peers to think of their narrow self interest Tenbrunsel and Messick, 1999 and lower trust (Mulder et al., 2006).

We implement four main treatments, giving a payoff of zero when exiting (T1), keeping only the contribution of the exiting peer as if she was working alone (T2), subsidizing it lightly (T3) or more heavily (T4). The prevalence of exit across treatments can be visually appreciated in Figure 2, in which darker colors indicate exit from the partnership.

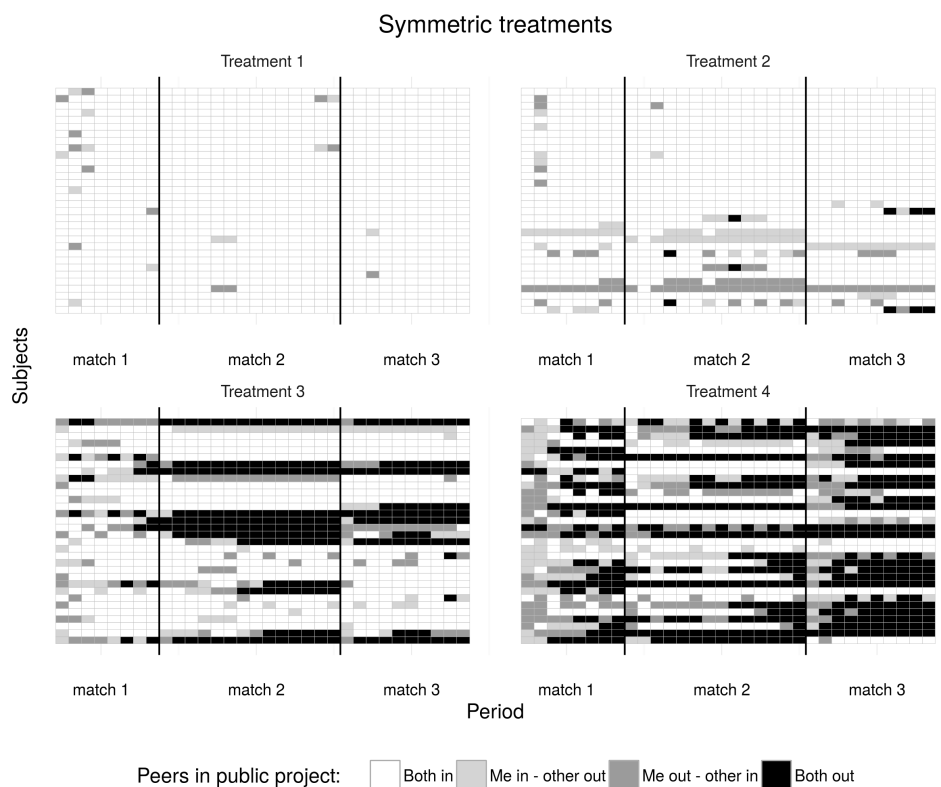


FIGURE 2: Prevalence of exit according to treatment in Gaudeul, Crosetto, and Riener, 2017

Figure 2 clearly shows, maybe not surprisingly, that exit is more likely the easier it is. But the welfare of the remaining dyads might be higher in treatments with high exit. We find instead that it is treatments with high barriers to exit that generate the highest welfare overall as they foster stability and prevent inefficient separation of pairs. There is excessive exit in treatments with low barriers to exit.

Since we elicit the beliefs of subjects (via a linear scoring rule) about both the likelihood of exit by the peer and her effort conditional on her being in the common or private project, we can distinguish two main drivers of exit: exit driven by cost-benefit considerations (the other is not contributing enough) and exit driven by a fear of being left alone in the project. We find latter to be more important when barriers to exit are low, less so when barriers are high, as already hypothesized by Hirschman, 1970.

Exit was not the only focus of my work. A paper with Werner Güth and Vittoria Levati from Jena, and Matteo Ploner and Luigi Mittone from Trento was instead focused on the *voice* component, investigating the role of voice over payoff inequality in punishment.

In Crosetto et al., 2012 we study a much more traditional sequential public good game played by a dyad. By the very nature of the sequential public good game with perfect information, a trusting first contributor can be exploited by a second contributor. We add a punishment phase, in which the first contributor plays again and can decide whether to punish the second contributor. We vary the modes and consequences of punishment to try to disentangle different motives of punishment, from voicing one's anger to reduction in payoff inequality. We have two treatments. The first implements the usual additive nature of punishment, in which the result of the punishment phase are simply added to (or subtracted from) payoffs of the public good phase. In such a treatment, punishment *both* voices one's anger *and* reduces payoff inequality. The second treatment implements an alternative payment scheme, whereby *either* the punishment phase *or* the public good phase are paid. In such a treatment, assuming subjects do not reason in expected terms, punishment can only *increase* inequality and does not reduce it. To explore the effect of "cooling-off" both treatments are run once with immediate and once with delayed punishment. Moreover all four treatments are also run in a *pure voice* condition with no monetary consequences of punishment whatsoever – punishment is in these treatment *virtual*, not carried out.

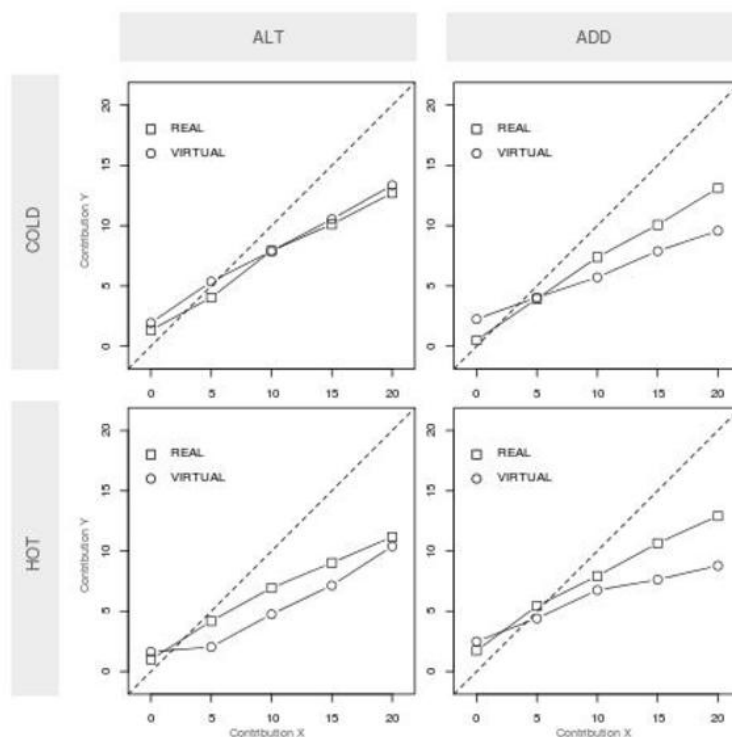


FIGURE 3: Mean public good contributions in Crosetto et al., 2012

Results show, across all conditions, the emergence of conditional cooperation. This result appears clearly in Figure 3 in which across all hot and cold, additive and alternative payment, virtual and real consequences, the contribution of the second player (Y) mimics closely, if at a discount, the contribution of the first player (X).

Punishment is generally low, but it is higher in the heat of the moment. Punishment is by

and large *not* used to reduce inequality, but rather to punish violators of the reciprocity norm, irrespective of the presence or not of monetary consequences, and irrespective of its inequality-reducing or -enhancing nature. Thus subjects punish more on grounds of principle, reacting to a norm violation, than by considering the possible consequences. Underlining the importance of voice, voicing one's anger or letting off steam are better descriptors of punishment in our experiment than any other consideration.

Entrepreneurial finance meets the Internet: crowdfunding

"How did you make all these people pay for music?" And the real answer is, I didn't make them. I asked them. And through the very act of asking people, I'd connected with them, and when you connect with them, people want to help you.

Amanda Palmer

Another area in which the accelerating effect of the Internet has been felt is entrepreneurial finance. Traditionally, firms have been funded through bank loans, the financial market, venture capital companies or angel investors. The possibility to reach wider and wider audiences and at the same time to cater to each and every *niche* market provided by the Internet opened up a new funding channel: crowdfunding. In crowdfunding, an entrepreneur – usually in the arts, communication, gaming industry, but not only – launches a call to fund her project. On specialized websites, individuals can donate, or lend money, or pre-purchase the product, or all of those together. Usually website use the threshold-pledge model: users pledge some money to the project (in exchange for a reward, a product, credits, or with nothing in return), and pledges are collected by the project launcher only if a certain threshold, predetermined by the project launcher, is met.

Crowdfunding has seen an incredible boom over the last five years. From being a niche business, it has reached center stage and some platforms, like Kickstarter, have collected more than 3 billion Euro of funds for over a hundred thousand successful projects. The magic worked by these websites is to connect start-ups, creative artists and social entrepreneurs with the people most likely to support them, wherever they might be located. But is this some unprecedented wave of altruistic behavior that the Internet can mobilize, or is this boom in project funding driven by other factors – rewards, social-image concerns, sponsorships, returns from investments?

Together with Tobias Regner we contacted Startnext, the biggest German crowdfunding platform, with over 45 millions raised for about 5 thousand successful projects. Startnext gave us something rare and precious: their whole database of transactions over all projects – successful or not. We have more than a hundred thousand pledges, and we know several characteristics of the projects and the rewards pledgers got in return. Our main interest in entering the literature was twofold. On the demand side, investigate the motives of pledges. Is there really a hidden reservoir of altruism to tap? Moreover, the data are perfect to study bandwagon effects. On the supply side, what are the main recipes for success?

This is the only non-experimental work of all those presented here. But there is a reason for that: with Tobias we hope to be able to find patterns in the data that shed light on the motives

of pledgers; and then design experiments to further disentangle the patterns we find. So far we have produced two empirical papers on our rich dataset, but more are to come.

In Crosetto and Regner, 2014, now under revision for Research Policy, we explore funding dynamics, the role of herd and bandwagon effect, and the possibility of self-pledges – project owners indulging in a limited form of insider trading and secretly funding their own projects. Self-pledges are banned from big American platforms, but they are allowed in Startnext, giving us a unique opportunity to see to what extent project launchers try to steer the crowd, and to what extent they succeed in doing it.

A key finding in the crowdfunding literature is path-dependence: pledge cascades usually happen at the very beginning of a project, and fast starts are the key predictor of a campaign's success (Colombo, Franzoni, and Rossi-Lamastra, 2015; Kuppuswamy and Bayus, 2015; Rijt et al., 2014). Our data tells us a richer story: while early cascades do exist and a fast start is a predictor of success, most of the eventually successful projects *underperform* the average project as late as at 75% of project time. Most projects are last-minute bloomers: this might be due to some (unobserved) form of network externalities that kick in once a project has reached a certain critical mass. Most information and pledge cascades happen during the final quarter of project duration. Moreover, as projects get nearer to the end of their allotted time to raise funds, the share of pledgers making free donations decreases, and the share of those just advance-buying the product increases. These findings cast some doubts onto the 'altruism' argument. They give some credibility to the claim that as the likelihood of a project being funded increases, a pledge becomes less of a commitment to help the project off the ground and more of an early purchase of a product.

The last-minute cascades could be driven or steered by self-pledges. By strategically engineering an increase in attention for a project, an entrepreneur could steer the crowd and secure funding. Indeed, we find evidence that project launchers *try* to do this, as self pledges come after weak periods, and are on average timed to get a cascade going. Nonetheless, there is no evidence that they *succeed*: we find no evidence of cascades triggered by self pledges.

In Regner and Crosetto, 2017 we investigate the role of social- and self-image in the selection of rewards. On crowdfunding platforms pledges are usually met with a reward. This can be a simple thank you, a product, a branded clothing item – t-shirt, baseball cap – an invitation to a special event, and so on. We restrict analysis to a sample of 14,031 pledges from 346 projects hosted by Startnext. We categorize reward levels based on the concept of identity (Tajfel and Turner, 1979), introduced by Akerlof and Kranton (2000) into economic modeling: behavior in line with one's identity results in positive utility, while behavior in contrasts with it has negative effects.

We hypothesize that such an identity/community feeling of crowdfunders can be supported by rewards that strengthen the relationship between the crowd and the project creator. Results show that providing more rewards that appeal to the crowd's identity increases the likelihood of project success. Specifically, success is correlated with the presence of rewards that offer active participation in the project. Our results are in line with the idea that a key to crowdfunding success is the entrepreneurs' ability to conjure a feeling of belonging to a group sharing a common cause, as underlined already by Belleflamme, Lambert, and Schwienbacher, 2014; Gerber and Hui, 2013.

Risk preferences

My main domain of research after the end of my PhD has been risk preferences. Risk preferences are an elusive concept. They cannot be directly observed, but are just a latent construct: they must be inferred from choices. The problem with inference from choices is that in order to obtain a measure of risk preferences you need both observation *and* a theory. Observations alone are not informative as long as you do not have a theoretical framework to use to make sense of them. Theory alone is not informative, for all its elegance, if it cannot be mapped to choice behavior and give clear predictions that are then borne out by observation.

In measuring risk preferences, researchers have to make several choices. First, they must assume that these preferences are (at least locally) stable. Then, they must decide if those preferences are a primitive of behavior, and apply to all domains alike; or they are domain-dependent, implying that a person can be very risk loving while in Las Vegas, and very risk averse when it comes to extreme sports or economic decisions. Then, short of having a theory of risk preferences evolution, they must assume that risk preferences are stable in time. All these steps have been dealt with by economists several times over; the basic model of *homo oeconomicus* assumes exogenous, fixed, stable, domain-agnostic risk preferences. Cognitive psychologists assume otherwise – and they have their own body of literature. In my works I have always chosen to stay firmly on the economists' side of the divide, and all works discussed here assume stable preferences at their core. Still, I have reasons to doubt this, and I am not alone in thinking that something is wrong in risk elicitation.

Assuming a single, stable, fixed risk preference is a first step in being able to identify it from choice data. A second step is the choice of the theoretical framework to be used. Here economists have developed several competing theories. Expected Utility Theory (EUT) posits that the utility of a lottery – a probability distribution over a set of events – is just the weighted average of the utility of each event, the weights being given by the probability of each event. Alternative theories make different assumptions about the way subjects process probabilities (Rank-Dependent Utility, RDU, Quiggin, 1982), or deal with losses (Cumulative Prospect Theory, PT, Tversky and Kahneman, 1992). Other theories exist and have been widely discussed and tested. Some of the most powerful tests pitched all theories one against the other, as did Hey and Orme, 1994.

But even when assuming a single theory – in this case, EUT – things can easily get confusing. Given the premises – fixed preferences, fixed theoretical framework – any set of lottery choices, wide and diverse enough to give some power, should be able to identify risk preferences. Different set of lottery choices might lead to different estimates only insofar as there is unmodeled or misspecified error or if the choice set happens to cover only a part of the possible space of choices. Unfortunately, this is not the case. Several contributions in the literature (Dave et al., 2010; Deck et al., 2013; Isaac and James, 2000; Menkhoff and Sakha, 2017) show varying degrees of inconsistencies across some of the most widely used tasks in experimental economics, even if all tasks are ultimately built of the same raw material – choices among lotteries – and all share the same foundations – fixed risk preferences and EUT.

My journey in the risk preferences literature, carried out jointly with Antonio Filippin, has taken three main turns. First, we have developed the Bomb Risk Elicitation Task, with the aim of improving on existing tasks. Second, we have entered the literature striving to compare risk

elicitation tasks, and find out the reasons for the often observed inconsistencies and low correlations. In doing so, we went from a general appraisal of different tasks, to more detailed studies investigating the role of several widespread lab practices about the resolution of uncertainty and the role of other-regarding concerns in the lab. Third, armed we had learned thus far, we have entered the debate on gender differences in risk preferences.

We now deal with these three steps in detail, before summing up what has been learned and what is in my opinion the road ahead.

The Bomb Risk Elicitation Task

There are too many risk elicitation tasks! Let's create yet another one!

anonymous

In order to elicit risk attitudes, economists have always relied on choices among lotteries. This comes directly from the theoretical definition of risk aversion as the curvature of the utility function. A lottery provides the most direct and straightforward test of the economic translation of the multi-faceted concept of risk.

Nevertheless, in experimental economics a lot of effort has been put into *a.* selecting the 'best' set of lotteries to give to the subjects, and *a.* packaging them into a *risk elicitation task* that could be easily understood by and administered to subjects, while being straightforward to analyze and giving consistent estimates of the risk attitude of individuals. The second step is by no means necessary, and several key authors in the risk literature prefer to give subjects several dozen lotteries and do not restrict their experiments to simple, pre-packaged tasks. Nonetheless, it is simple and compact tasks that are the most widely used approach, especially for economists that are not specialists of risk elicitation but elicit risk in other contexts. It is hence important that these tasks deliver precise, unbiased, and reliable estimates.

There are several competing tasks that have been proposed in the literature to elicit risk preferences. While they are all pre-packaged variations of choices among lotteries, they differ in the number of choices, in the domain they span (risk loving, risk aversion, or both), in their frame (plain, as an investment decision, as a gamble, as a game...), in their length and complexity.

I never thought that the existing risk elicitation task were very good at their job. They are clever designs proposed by clever people, but most of them fail in one respect or another. Most tasks are difficult to understand. One particularly well known case is the Holt and Laury, 2002 task, in which $\sim 15\%$ of subjects submit choices that are inconsistent, impossible to rationalize, and must routinely be discarded from the analysis. Some tasks, like the standard implementations of the tasks proposed by Eckel and Grossman, 2008b and Gneezy and Potters, 1997, span only the risk averse domain, imposing on subjects the unproved assumption that risk lovers do not exist or are extremely rare. Some other tasks give only coarse results, binning subjects in a handful of categories, while a more precise estimation might be useful. Moreover, despite the fact that risk is a matter of interest to both economists and psychologists, there is a deep and vast divide, conceptual and practical, among tasks used by psychologists – who use deception, prefer real-world scenarios and stimuli, often hide probabilities and use visual devices – and

economists – who value abstract designs, use tables of numbers, and prefer choices that can be easily fed to their mathematical models.

In introducing the Bomb Risk Elicitation Task, Antonio and I wanted to create a task that might overcome some of the weaknesses of the existing task, and bridge the gap between economics and psychology. We needed a task with several choices in order to yield finer results, but that would still be simple and not time consuming. We needed a task exploiting visual cues rather than just an arid mathematical notation, but that kept the rigor needed by economic theory. We needed a task that would naturally span both the risk aversion and risk loving domains, without imposing a high cognitive burden on subjects. A task that might be easily deployed in the field but that would do well in the lab as well.

We came up with the Bomb Risk Elicitation Task (BRET), first published in the Journal of Risk and Uncertainty in 2013 (Crosetto and Filippin, 2013).

In the BRET subjects face a 10×10 square in which each numbered cell represents a box. They are told that 99 boxes are empty, while one contains a time bomb programmed to explode at the end of the task, i.e., *after* choices have been made. The position of the time bomb $b \in \{1, 100\}$ is uniformly random determined after the choice is made. Calling k the number of boxes collected, if $k_i^* \geq b$ it means that subject i collected the bomb, which by exploding wipes out the subject's earnings. In contrast, if $k_i^* < b$, subject i receives 10 euro cents for every box collected. The metaphor of the time bomb allows us to observe any choice, avoiding the truncation of the data that would happen in case of a real-time notification.

Subjects' decision can be formalized as the choice of their favorite among the set \mathcal{L} of 101 lotteries fully described both in terms of probabilities and outcomes by the stopping point $k \in \{1, 100\}$,

$$\mathcal{L} = \left\{ \begin{array}{ll} 0 & \text{with probability } \frac{k}{100} \\ 0.1k & \text{with probability } \frac{100-k}{100} \end{array} \right. .$$

k drives at the same time probabilities and gains, summarizing the trade-off between what can be earned and the likelihood of earning it. The degree of risk aversion negatively correlates with the choice of k and a risk-neutral subject should choose $k = 50$. The visual representation of the BRET after 45 seconds and with a worth of 10 cents per block is given in Figure 4.

In creating the BRET we did not start from scratch, but, as always, we were standing on the shoulders of giants. The BRET is similar in spirit to the Balloon Analog Risk Task (BART, Lejuez et al., 2002), in which a subject must repeatedly inflate a balloon that risks bursting at any time. The BART is in turn similar to the Devil's task (Slovic, 1966), a risk task to be used with children, in which subjects turn cards and earn points for each card, but lose all in case a devil card is turned. There are substantial and important differences, though. Differently from these other tasks, the BRET is designed to display probabilities in a clear, precise yet intuitive way and to avoid any truncation of the data. Being composed of 100 boxes, probabilities appear both as numbers (out of 100) and visually as areas. This display of probabilities does not hide them (as both the Balloon and Devil's task do) nor requires any computation on the part of subjects. Relying on a *time bomb* metaphor, the BRET delays the moment of the "explosion" to after the

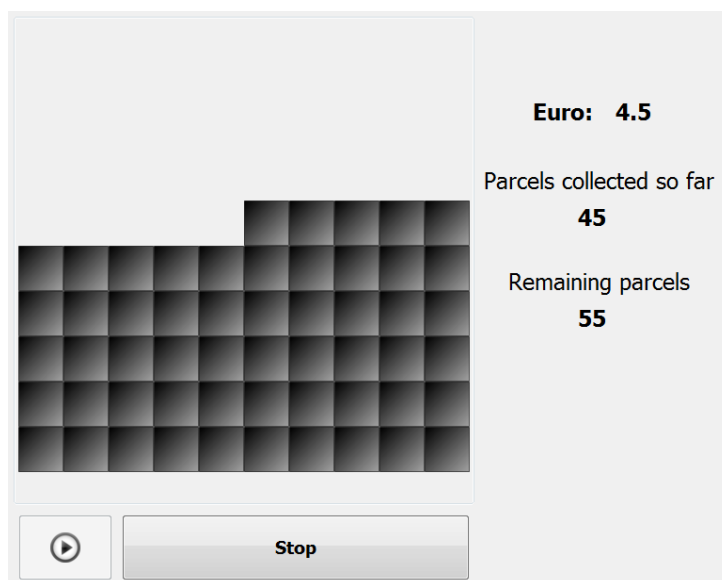


FIGURE 4: The standard interface of the Bomb Risk Elicitation Task

choices are made, and thus avoids any truncation of the data. Irrespective of the degree of risk loving, all preferences can be observed. This is not the case for neither the Balloon nor the Devil's task, in which subjects can burst the balloon or turn the Devil's card; researchers can only observe the minimum between the subject's preferences and the truncation due to the random device used. This truncation is far from neutral, since it affects results asymmetrically: the more a subject loves risk, the more we are likely *not* to observe her true preferences.

The Crosetto and Filippin, 2013 paper introduced the BRET theoretically, showing what is going on under the hood, showed result of a first run with 269 subjects, and provided a large set of experimental robustness checks.

Figure 5 shows the main empirical characteristics of the BRET. The task yields a rather continuous measure, skewed towards risk aversion; the modal choice is very close to risk neutrality (with a light shade of risk aversion); and there is a not negligible share of subjects that display risk loving preferences.

To check the robustness of these results to different dimensions of the BRET, we ran several experimental tests, involving altogether 1255 subjects. The BRET was run without its dynamic aspect, with live explosion (like the Balloon), without a trial period, with higher stakes, with more and less blocks, with faster deletion time, with random rather than sequential collection of blocks, with a loss frame, with wealth effects (i.e., after a previous task whose payoff was known), repeated rather than one shot. The BRET proved altogether robust to these changes. Choices reacted significantly only to increased stakes – in line with the literature, subjects appear more risk averse when faced with higher stakes – and to wealth effects – with a convex U-shaped relationship between previous gains and risk tolerance.

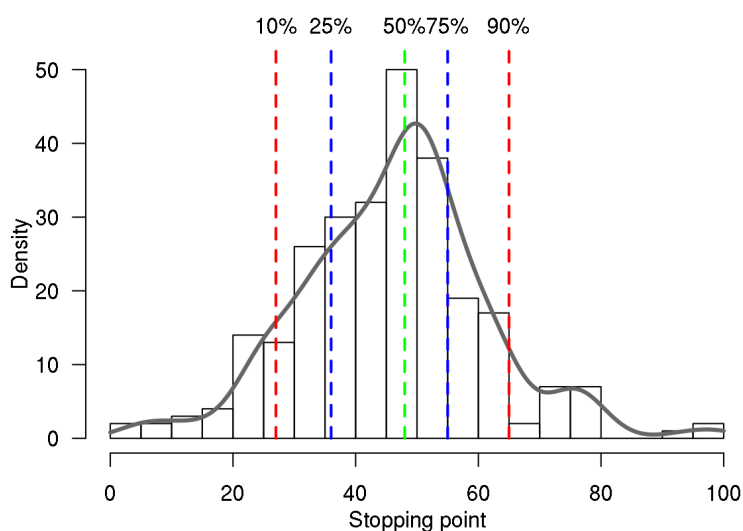


FIGURE 5: results of the Baseline BRET in Crosetto and Filippin, 2013

Why is there no consistency across Risk Elicitation Tasks?

I was going to thrash them within an inch of their lives, but I did not have a tape measure.

Groucho Marx

Having developed the BRET and tested its robustness, the most important thing left to do was to compare it with other risk elicitation tasks. Both me and Antonio were naive enough to think that comparing different tools would be an easy endeavor: how difficult can it be to run a handful of between-subjects experiments and then report the bare results or a structural estimation? But, as we discovered along the way, risk elicitation tasks hide behind their apparent simplicity a great deal of intricacies, and comparing them, far from being an easy task, it is a very tricky, and sometimes even impossible, exercise.

We started with a simple comparison, carried out already in the original BRET paper (Crosetto and Filippin, 2013). We ran a mix of between- and within-subjects sessions, and reported some simple summary statistics. But, as it is more often than not the case with risk elicitation task comparisons, we realized that the estimates given by each task varied widely.

Figure 6 gives the kernel density estimation of the results of our between-subjects sessions (~ 90 subjects per task). The figure reports results transformed into risk aversion coefficient r of a CRRA power function $U(x) = x^r$ for each of four risk elicitation tasks: Holt and Laury – HL, Eckel and Grossmann – EG, Gneezy and Potters – GP, and the BRET.

Even taking into account that these are between-subjects estimations – so that extremely risk averse subjects might have ended up participating in the EG task and risk lovers in the BRET – the differences are noticeable. They are also important in economic terms: they imply that the average subject in the BRET would have a certainty equivalent of 45.8 € for a 50/50 lottery

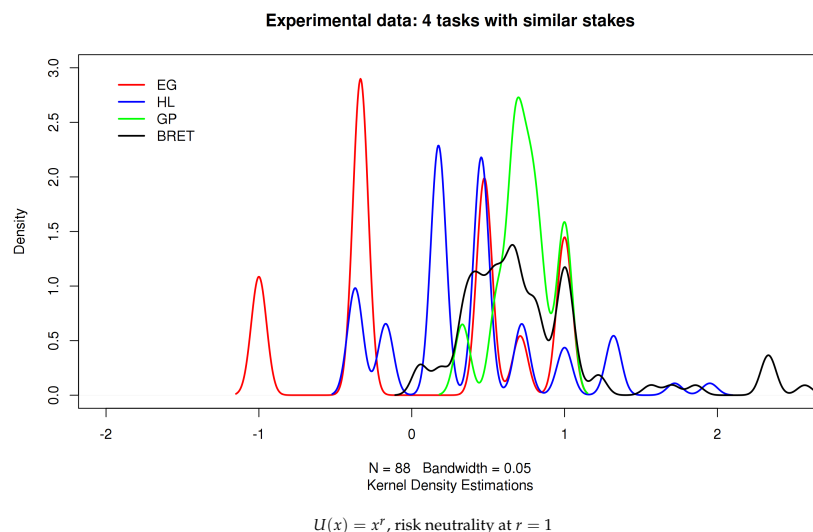


FIGURE 6: Estimates of CRRA risk aversion coefficient from different tasks

between 10 and 90 €; the average subjects in the EG task would be willing to pay only 31.6 € (37.3 for HL, 44.9 for GP). The same subjects proved not significantly different when self-reporting their risk attitudes. It looks as if subjects that self-report to be indistinguishable in terms of risk attitudes instantly reveal themselves as very different once looked at through the lenses of different tasks. Looking at the literature, it routinely appears that this is the case: estimates collected through different tasks are hardly comparable (Bruner, 2009; Dave et al., 2010; Deck et al., 2013; Harbaugh, Krause, and Vesterlund, 2010; Isaac and James, 2000; Menkhoff and Sakha, 2017; Reynaud and Couture, 2012). Why?

There are (at least) two competing perspectives from which to look at this issue. The one that is usually preferred by experimental economists is that you cannot trust self-reported and not incentivized answers. If this is the case, the straightforward interpretation is that the tasks made existing, but hidden differences emerge. The competing perspective is that subjects are indeed similar in their risk attitudes, and the tasks introduce each one a task-specific bias in the observed choices. If this is the case, it becomes extremely important to know a. which tasks impose which bias; b. if this can be predicted and corrected for; and c. if there is a task that is bias-free, or at least reduces it.

The bias induced by the tasks could be due to the characteristics of the tasks themselves or to other, at first sight irrelevant, details of the sessions. We decided to look into this issue in detail, and the results came out in three different papers. Crosetto and Filippin, 2015 dealt with the intrinsic characteristics of the tasks. Filippin and Crosetto, 2016b investigated the role of the randomization device used, checking if giving the subject more control (i.e., by letting them roll the dice) induced *illusion of control* and higher risk taking. Crosetto and Filippin, 2017b dealt with two further possible issues: the sound cues given away by clicks during a session and the role of a common rather than individual resolution of uncertainty.

Here are all the spoilers in a few lines: there is a task-specific bias; most of it is due to the mapping of choices into risk aversion coefficients, or in other words, to the choice of the lotteries given to subjects; nonetheless, the differences are not reducible to this factor only; in line

with most experimental economists we found no role for illusion of control; and sound cues from other subjects do play a role, but only when the resolution of uncertainty is also common, hinting at an inequality aversion explanation.

For the interested reader, I will give some details in the following paragraphs.

"Structural" task-specific bias: the characteristics of the tasks

In Crosetto and Filippin, 2015 we explored how much of the difference in estimates across task can be attributed to the structural characteristics of the tasks themselves. We did so in two distinct ways: theoretically, and using a set of simulations.

In our theoretical argument, we showed how the tasks, *via* the lotteries they choose to give subjects, explore only certain parts of the mean-variance space, and hence give estimates distorted by construction.

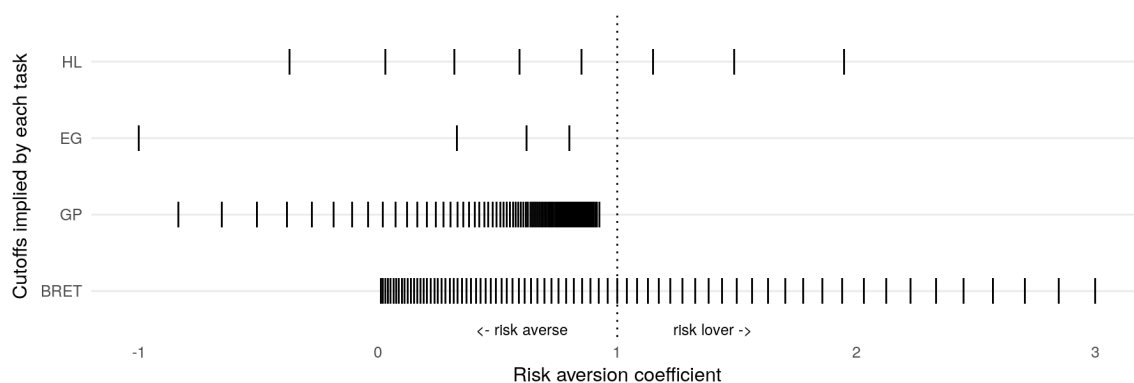


FIGURE 7: Mapping of choices in each task to CRRA x^r coefficients

The intuition for this finding is shown in Figure 7. Each vertical line indicates which risk aversion coefficient is assigned to the subject making each possible choice in the task. Three things jump to the eye: the tasks span only subsets of the parameter space; some tasks give rather sparse intervals; and the intervals between vertical lines are, excluding HL, not constant, implying a nonlinear mapping between choices and risk aversion parameters. To make an example, a subject whose true r were 0.75 – a mildly risk averse subject – would make 5 safe choices in HL, would choose the riskiest available option in EG, nearly the riskiest available in GP, and would collect about 35 boxes out of 100 in the BRET.

This fact has three order of consequences, depending if we assume deterministic or partly stochastic preferences, and, within the latter, if we assume the error to be white noise or to be subject to framing effects. In order to show these consequences we ran three sets of simulations. We generated a population of 100.000 virtual subjects characterised by $U(x) = x^r$, $r \sim N(0.8, 0.3)$. We exposed each virtual subject to each of the tasks, collected their choices, and then analyzed the data, retrieving the r implied by agent's choices and running some descriptive statistics on the data.

If we assume that subjects have completely deterministic preferences, the bias imposed by each task is minimised but still not absent. Since some tasks (as EG and GP) cannot tell apart risk

lovers from risk neutrals, average estimates from the task will be downward biased – the more so, the more risk lovers exist in the population. Moreover, the coarseness of some tasks is an additional source of bias. EG estimates give a mean of 0.743 (true mean: 0.8), GP estimates 0.756. This means that even in the best possible scenario – in which we know and control deterministic preferences – two tasks impose a significant downward bias on the estimates.

If we assume preferences to be stochastic, the bias imposed by each task increases. We first modeled the error around the true preference to be white noise – i.e., subjects have an r , but behave as if they had $\hat{r} = r + \varepsilon, \varepsilon \sim N(0, \mu)$. This means that the error is *not* task-specific. Subjects are not subject to framing effects; rather, they just have an unbiased tremble around their true preference. If we set a moderate variance of the white noise to $\mu = 0.6$, the bias induced by the tasks become noticeable. EG estimates give a mean of 0.562 (true mean 0.8), GP gives 0.623, the BRET gives 0.84, and only HL is robust to noise and gives a really good 0.804.

But there are plenty of reasons to think that errors can be task-specific. We know for instance that people are subject to framing and context effects. For the sake of illustration, Figure 8 shows what would happen if subjects were heavily influenced by framing effects, and people rating themselves as very risk loving simply chose the riskiest option available in each task, subjects rating themselves as extremely risk averse the less risky option, and so on. That is: subjects do not choose on an absolute scale, but rather on the scale given to them by the task.

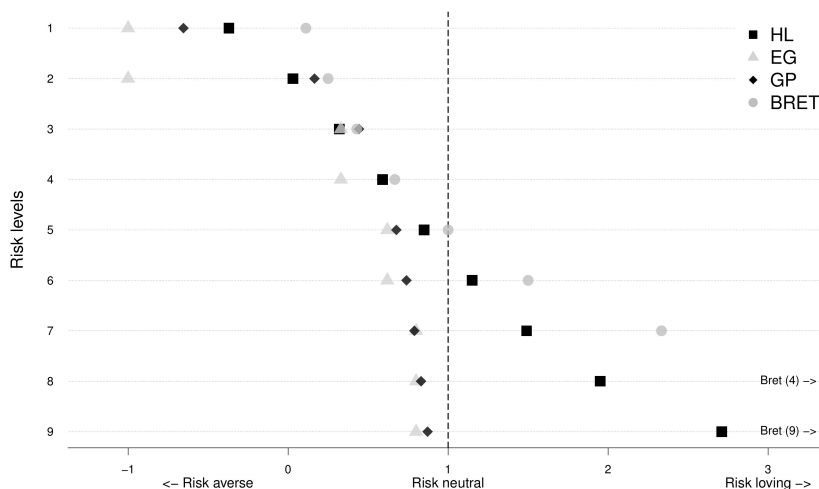


FIGURE 8: Mapping from choices to r parameter assuming framing effects

This is admittedly an extreme assumption, but it allows to give the intuition in the clearest possible way. Figure 8 shows that an extremely risk averse person can be assigned r parameters from -1 to 0.2, according to the task; conversely, a very risk loving person is estimated to be a risk averter by two tasks, is assigned an r of 2.6 by HL, and a very high r of 9 by the BRET. If we allow for framing effects, the biases introduced by the tasks increase dramatically.

Our simulations confirmed this intuition. Imposing deterministic preferences but allowing just 10% of the subjects to randomly choose in the task space – that is, error is not white noise but task dependent, a *very* mild form of framing effect – resulted in estimates going *more* off the

mark than with much more volatile white noise. EG estimates gave 0.68, GP 0.53, the BRET 1.12, and again only HL performed well, giving 0.811 (true mean: 0.8).

The bottom line is that *any* risk elicitation task imposes, *by construction* a bias on the observed choices. Does this explain *all* of the across-task heterogeneity of risk preference estimates? No. Yet, a significant part of it can be accounted for by this usually overlooked structural characteristic. It looks as if tasks are colored glasses, and each colors the world we see with its own peculiar hue.

The influence of randomization devices, sound cues and common fate

For want of a needle, the kingdom was lost.

Traditional proverb

The previous results showed that measuring risk attitudes is a tricky endeavor; biases are introduced with each and every measurement tool used. Risk attitudes are a latent construct, and they can only be imperfectly observed and measured, usually in conjunction with other personality traits or other preferences. It is not surprising then that other, at first sight minor, characteristics of risk elicitation tasks end up influencing the estimates collected in laboratory sessions.

Together with Antonio we used the large amount of observations gathered with the BRET to shed some light on three possible sources of bias: the illusion of control derived from having a more active role in the choice process and in the resolution of uncertainty; weak sensory interactions among subjects, namely sound cues given away by subjects by clicking or hitting their keyboard in the lab; and common *vs.* individual fate, i.e., the fact of being subject to one common rather than several individual random shocks.

In Filippin and Crosetto, 2016b we dealt with illusion of control – the belief to have control over pure chance events. The intuition is simple. In real life, performance in tasks is usually a convex combination of skills and luck. Gambling tasks, lotteries, and, incidentally, risk elicitation tasks are pure luck situations, in which skills play no role. Literature from psychology Benassi, Sweeney, and Drevno, 1979; Langer, 1975; Presson and Benassi, 1996 repeatedly found that giving the subjects an active role in the pure luck situations – by, e.g., allowing them to choose the lottery ticket rather than being assigned one, or letting them roll the die – enhanced the beliefs of success, as a consequence resulting in higher risk tolerance.

Experimental economists repeatedly failed to replicate the finding in incentivized economic experiments (Charness and Gneezy, 2010; Poon, 2011), or did find great individual heterogeneity (Li, 2011). Interestingly, experimental economists always induced illusion of control by manipulating the degree of involvement of the subject in the *resolution of uncertainty*. This was nearly always done by letting the subjects roll the die themselves. In the psychology literature, another form of control is used: control over the *choice of the lottery*. For example, Langer, 1975 ran a lottery with football players cards as tickets, and let subjects choose their favorite player.

In our paper we implemented both manipulations, using the BRET. We replicated the literature by letting subjects roll two ten-sided dice to determine the position of the bomb. We gave further control to subjects by giving them the dice at the beginning of the session and letting

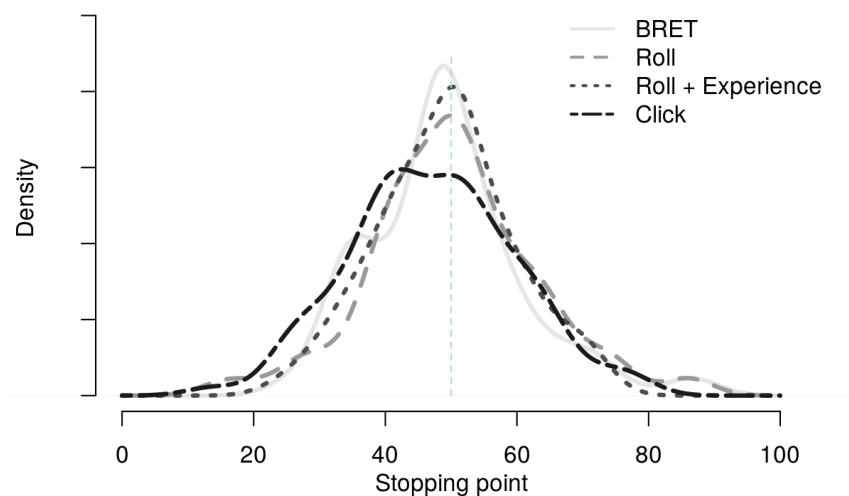


FIGURE 9

them ‘warm up’ by rolling them as much as they wanted before the beginning of the session. But we also gave subject more control over the *choice* of the lottery, by letting subjects actively click to collect boxes – rather than just wait for them to be collected by the computer, as in the standard BRET.

We find null results across the board. Neither the perceived probability of winning (not shown) nor the choice made (Figure 9) react to the experimental manipulations. Our results rule out that illusion of control might be triggered by more control over the choice of the lottery – in particular in the BRET. The reason to the discrepancy between economic and psychology results is still to be found; our paper indirectly points in the direction of monetary incentives.

In Crosetto and Filippin, 2017b we investigated the role of conformism in risk elicitation, exploring two possible channels through which it can impact choices: through sound cues that give away a fuzzy signal of other subjects’ choices; and through the ex-post commonality in payoffs induced by a common, rather than individual, random shock.

Broadly speaking, there are two orders of arguments raised to rationalize the conformism often observed in the lab, either by psychologists (Asch, 1955; Asch, 1951) or economists. Conformity might arise due to preferences – an intrinsic taste in adapting one’s behavior to follow an external reference (Haney, Banks, and Zimbardo, 1973), or a taste to comply with social norms (Bicchieri and Xiao, 2009); or it might be a consequence of information search – there is information to be gathered from other subject’s behavior, absent any specific preference to act like others (Anderson and Holt, 1997; Banerjee, 1992). A third, less explored idea refers to inequality aversion: on the face of common random shocks, conforming to others ensures that ex-post payoff differences will be minimized. Economists have traditionally held to the information-driven explanation – for instance according to the theory of herd behavior and information cascades (Banerjee, 1992). Nonetheless, experimental evidence shows that conformity arises even in context in which a private and public signal are given, and the public signal is not informative and *known* by the subjects to be so (Goeree and Yariv, 2015).

In this paper we use the “Click” version of the BRET, the same used for the illusion of control treatment, exploiting the fact that subjects can hear each other’s clicks. These clicks indirectly give away a fuzzy signal of the other subjects’ willingness to take risk. Manipulating the audibility of the signal by making the subjects use (or not) earmuffs, we are able to pin down the effect of the signal on subject’s choices. This allows us to do two things: first, to ascertain if sound cues can have an effect on the measure of risk (and other) preferences; and second, to contribute to the literature on conformism.

	(1)			(2)		
	Stopping point			Stopping point		
<i>Constant</i>	56.76	**	(2.31)	51.61	**	(2.12)
<i>Idiosyncratic Risk</i>	2.991		(1.47)	1.674		(0.61)
<i>Correlated Risk</i>	7.901	***	(3.62)	-0.302		(-0.10)
<i>SOEP</i>	1.867	***	(4.24)	2.015	***	(4.68)
<i>female</i>	3.138	*	(1.95)	-2.084		(-0.79)
<i>age</i>	-1.155		(-0.86)	-0.737		(-0.54)
<i>age²</i>	0.0299		(1.25)	0.0216		(0.88)
<i>complexity</i>	0.115		(0.36)	0.244		(0.79)
<i>Nsubjects</i>	-0.352		(-0.57)	-0.282		(-0.47)
<i>Smaller lab</i>	-5.876		(-0.74)	-4.900		(-0.64)
<i>Idiosyncratic Risk</i> × <i>female</i>				1.986		(0.60)
<i>Correlated Risk</i> × <i>female</i>				13.62	***	(3.40)
<i>N</i>	267			267		

t statistics in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 1: Multivariate regression analysis

Our findings are given in Table 1. In the table, the *Isolated* treatment had subjects wear earmuffs, and an individual resolution of uncertainty. *Idiosyncratic risk* removed the earmuffs; *Correlated risk* had no earmuffs and a common resolution of uncertainty. Treatments were run between subjects. We observed a slight (weakly significant) shift towards more risk taking when removing earmuffs. When adding the common resolution of uncertainty, the shift towards risk taking is impressive, and the average subject is a moderate risk lover. The regression shows that all of this quite dramatic shift towards estimated risk loving is driven by women only.

In a nutshell, hearing the others click leads to conformist behavior and – given the nature of clicks in the BRET – to more risk taking, but only for women and only in a situation of correlated risk and common fate. This is consistent with earlier evidence showing that women tend to be more susceptible to social influence than men, mainly in public interactions or in group pressure situations (Eagly and Carli, 1981).

Which leads us to the role of gender in risk attitudes, which we will discuss next.

Is there a gender difference in risk attitudes?

Men are from Mars, Women are from Venus

John Gray

The literature on gender differences in psychology and economics is vast and varied. Gender differences have been documented in competition (Gneezy, Leonard, and List, 2009; Gneezy, Niederle, and Rustichini, 2003), in overconfidence Barber and Odean, 2001, and also in risk preferences. Women are widely believed to be more risk averse than men, as confirmed by several reviews of experimental economics Charness and Gneezy, 2012; Croson and Gneezy, 2009; Eckel and Grossman, 2008a.

This finding is so entrenched in the mind of experimental economists that most treat it as a *fact* that can be used to explain, for instance, gender wage differentials. It is widely considered one of the few robust stylized facts emerged from experimental economics, and failure to replicate it makes researchers doubt their samples (as done, for instance, by Anderson and Mellor, 2009).

Mine and Antonio's interest in this issue came directly from a similar doubt: despite gathering thousands of observations, we found *no* gender differences in the BRET. At first we thought something might be wrong with our task. This prompted us to look around for gender results in other tasks. An in-depth look at the literature resulted in two main, shocking findings: gender differences in risk attitudes are less ubiquitous than they are depicted to be, and the likelihood of observing gender differences correlates with the task used.

Our survey of the existing literature already made clear that in experimental economics gender scholars had always focused on a limited set of tasks, most notably the Eckel and Grossman, 2008b and the Investment Game by Gneezy and Potters, 1997. When looking at those tasks only, evidence for gender differences seems indeed robust (despite some findings being disputed on statistical grounds, Nelson, 2013, 2014). But a survey of about 150 studies using a broad definition of risk, from smoking to driving to gambling, and analyzing self-reported, incentivized, as well as observed choices (Byrnes, Miller, and Schafer, 1999) found gender differences that were usually small and significant only about half of the time. Moreover, the most used and most widely cited risk elicitation task in experimental economics, Holt and Laury, 2002 (HL), had not been subject to a close look from a gender perspective. We decided to do it.

Gender differences and outcome reporting bias

Our first step was to gather all papers citing HL, about 550. Out of those, only roughly a fifth had made an experimental replication; only 20 papers reported results by gender; and only 3 found significant gender differences. This is in stark contrast to results obtained with other tasks.

At this point we faced a first issue. More than 80 papers had run a HL replications *without* reporting results by gender. Why? On the one hand, this could be due to the simple fact that risk attitudes are usually thrown in the paper as controls, and they are not the main focus of the paper. Reporting them is somewhat not required, and reporting a breakdown by gender even less so. But there is another, less benign interpretation: given the fact that researchers

expect to observe gender differences, finding a null result is somewhat suspicious; and absence of differences might be hidden to avoid worrying the referee about the nature of the sample. In other words, we might have been facing a case of *outcome reporting bias* – outcomes not in line with one's expectations are selectively not reported.

Together with Janna Heider, our research assistant in Jena, we gathered and read the 528 papers citing HL at the time, then contacted the authors of each study and asked to have access to their data. A majority of authors responded positively to our inquiry, allowing us to piece together a large dataset covering 54 published studies, for a total of 7654 subjects. Given the data, we could perform a simple analysis: we knew if a result was there or not, and we knew if it had been reported or not. We would have had evidence of outcome reporting bias if the likelihood of reporting gender results correlates with findings in line with expectations.

Results are described in Crosetto, Filippin, and Heider, 2015. We found no evidence of outcome reporting bias. The only significant predictors of a full report of gender results were the role of risk in the paper – results are more likely to be reported if the paper is about risk than if risk is just a control – and, interestingly, the share of women among authors. This was definitely good news for the experimental economics community, since even in presence of a strong prior consensus on a result, authors finding null results did not appear to engage in hiding the not-expected results on purpose; mostly, the non-reporting was due to the peripheral nature of risk and gender in the paper.

Gender differences in the Holt and Laury task

Once cleared the profession of scientific misconduct charges, we could turn our attention to gender differences in HL. In Filippin and Crosetto, 2016a we cleaned and brought the data to a common format, and we ran exactly the same statistical analysis – ranksum tests and t-tests – on each paper. Results of the ranksum tests are graphically depicted on Figure 10. In 11 papers out of 54 women were estimated to be willing to take *more* risk than males, even if never significantly so. In the other 43 papers, women appeared more risk averse than males, but this result was significant (at 5%) in 5 papers only, less than 10% of all papers. This result was independent of sample size, and replicated with t-tests and using paper-by-paper structural estimation by maximum likelihood including different error structures. Gender differences appeared to be the exception rather than the rule when using HL.

Statistical significance is nonetheless not the only important element to be taken into consideration. Effect size, especially when dealing with measurement issues, as I strive to do, is key. Unfortunately in economics, even in experimental economics, there is not a tradition of using effect size metrics. They are not generally taught, so young economists grow up oblivious to them; they are seldom used and not routinely demanded by referees. I didn't know much about them until the need for them arose in this project. The question is: given that an effect is found – here, gender differences – how big is it? This is of paramount importance in applied work, because further use of the base finding in applied and derivative work – e.g. investigating the reasons for a gender wage gap – depends on the magnitude, not just the existence of an effect.

We computed Cohen's d – a sample-size-independent measure of effect size – for each paper. Cohen, 1988 gave rule of thumb thresholds for interpreting its magnitude: up to 0.5 is a *small* effect, then a *medium*, and starting from 0.8 a *large* effect. Even *large* must be interpreted with

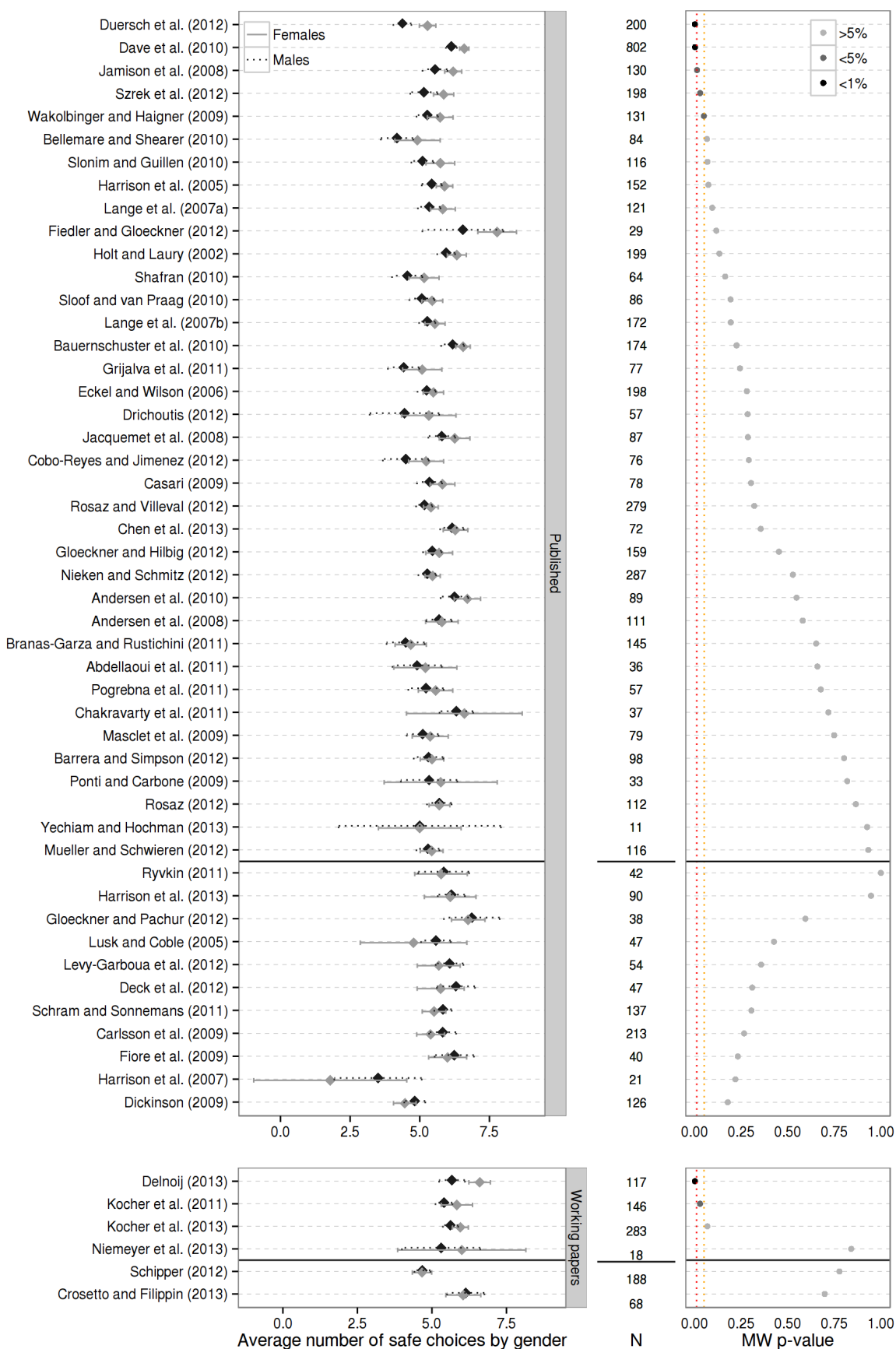


FIGURE 10: Gender gap in risk attitudes and its significance across papers

a grain of salt: in order for an effect to be interpretable at the individual level level – i.e., predicting with high accuracy a subject’s gender observing his or her risk aversion only – a d of 2 or more is needed, with a value of 4 meaning almost absolute discriminatory power (Nelson, 2014). Applying these threshold to our dataset only 3 papers find a medium effect, the rest being no or small effects.

These results might be due to the limited size of the samples used in experiments. HL might just be a noisier measure of risk attitudes, and hence gender differences might not be detected at lower sample sizes. To tackle this issue, we took two complementary roads. First, we showed that *after excluding inconsistent choices*, which most researchers do, HL is *not* noisier than other tasks, but actually less noisy. This resonates with our findings when dealing with the structural characteristics of the tasks, in which HL proved robust to different types of noise, while other tasks introduced sizable biases. Second, we pooled the data from all papers, thus analyzing a very large (from experimental standards) sample of more than 7000 observations.

In the pooled sample, gender differences are indeed significant, under a variety of estimation techniques and also under several functional form and error structure assumption in a maximum likelihood structural model. Nonetheless, they are very small in magnitude. The Cohen’s d of the pooled sample is a tiny 0.17, a sixth of a standard deviation. To give an example of how small this is, consider that if we compared two randomly drawn individuals from our sample, the most risk averse of the two would be a woman only 54.78% of the time, against a 50% chance if the two subpopulations were indistinguishable. This is a very tiny margin on which to build narratives assigning differentiated social outcomes by gender to the risk tolerance gap.

The role of safe options in gender differences

After gathering all this evidence, it seemed clear to us that the likelihood of observing gender differences in risk attitude was task-dependent. But why? Which characteristics of the tasks could trigger differentiated behavior by gender?

After reviewing the literature, we found out that robust observation of gender differences correlates with a. the availability of a safe option among the set of alternatives, and b. the use of 50/50 lotteries rather than other probabilities. These two characteristics unfortunately perfectly correlate with each other in existing tasks, so that it is impossible to disentangle the role played by each of them. As a consequence, we decided to run an original experiment, by directly manipulating the tasks.

In a recent working paper (Crosetto and Filippin, 2017a) we focused on the role of safe options in triggering gender differences. We focused on several rather than just one task, since we firmly believed and proved in previous work that tasks introduce task-specific biases. As a consequence, evidence collected with just one task is a composition effect of the manipulation applied to the task *and* its specific characteristic and biases. We manipulated three tasks – HL, BRET, EG – introducing a safe option where it was not available (HL and BRET) and removing it where it was (EG). In the BRET, this was done by guaranteeing the subjects that the bomb would *not* lie in boxes 1 through 25. This means that by stopping at 25 boxes, subjects get an amount of money with probability 1. In HL, we transformed the safe lottery into a safe amount. In the EG, we did the opposite, and transformed the degenerate lottery that is usually included in this task into a proper lottery, differentiating the two amounts.

	(1) BRET	(2) BRETSafe	(3) HL	(4) HLsafe	(5) EGnosafe	(6) EGsafe
ρ	0.21*** (0.06)	0.049 (0.07)	0.72*** (0.1)	0.67*** (0.08)	0.45*** (0.09)	0.46*** (0.08)
ρ_{female}	-0.024 (0.07)	0.13* (0.08)	0.019 (0.07)	0.12** (0.06)	0.12** (0.06)	0.20*** (0.05)
ρ_{SOEP}	-0.011 (0.008)	-0.0086 (0.007)	-0.049*** (0.02)	-0.044*** (0.01)	-0.028** (0.01)	-0.030** (0.01)
μ	0.16*** (0.03)	0.33*** (0.06)	0.67*** (0.1)	0.60*** (0.07)	0.27*** (0.05)	0.36*** (0.06)
μ_{female}	0.036 (0.05)	-0.12* (0.07)	0.059 (0.1)	-0.13 (0.09)	0.027 (0.08)	0.098 (0.09)
μ_{easy}	0.00065 (0.004)	-0.00035 (0.006)	0.059* (0.04)	0.051* (0.03)	0.029* (0.02)	-0.012 (0.02)
N	208	191	164	164	134	145
Log likelihood	-5274	-4132	-545	-425	-296	-321

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 2: Maximum likelihood structural estimations, Fechner error specification

We re-used some data from previous experiments, and we ran sessions involving altogether 1085 subjects, in a pure between-subjects experiment. The main goal was to see if we could pin down the appearance of gender differences to the presence of a safe option; and, if so, if we could give some hints as per the possible theoretical foundations of this fact.

Results show that the presence of a riskless alternative does not entirely explain the gender gap but it has a significant effect in triggering or magnifying (when already present) such differences. The main regression result is shown in Table 2. The Table reports the results of a structural estimation done with maximum likelihood, assuming a CRRA utility function and a Fechner error specification. Gender differences, absent in the standard BRET and HL, do significantly appear once we introduce safe options. On the other hand, removing a safe option from EG does not make gender differences disappear. Nonetheless, the change in the magnitude of the gender risk coefficient ρ_{female} is comparable across tasks, at ~ 0.1 . A safe option does not explain all the difference in observed risk attitudes – we did not expect it to do so – but there seems to be very solid evidence that it does contribute to make female more risk averse than males. This rather stable effect of the safe option in explaining gender differences across task is particularly remarkable, given the fact that risk elicitation measures are very volatile and task-dependent.

Why do safe option induce female to become more risk averse? A proper test of this question requires to purposely build a menu of lotteries capable of identifying the different determinants – a thing we did not do so far. Nonetheless, our results suggest that loss aversion and certainty effects likely play a role. Females may take less risks when there is actually something at stake that can be lost and they display a stronger inclination to avoid any risk when given the possibility to do so.

Context effects

One of the main pillars of Expected Utility Theory (EUT) is the Independence from Irrelevant Alternatives (IIA) Axiom. It states that if an alternative X is preferred to an alternative Y , then introducing a third option Z should by no means make Y preferred to X . Several violations of this axiom have been documented in time, dating back to the Allais Paradox (Allais, 1953). The identification, documentation and measure of these violations has been one of the core research programs of experimental and behavioral economics, and several dozens such violations have been documented (for instance, see the series about *anomalies* of the Journal of Economic Perspectives).

A group of IIA violations has been dubbed *context effects* (Tversky, 1972; Tversky and Simonson, 1993), since they play out when manipulating the choice sets – the *context* in which a certain alternative is chosen. For similar reasons (if for different methodological aims) such effects have been seen as the result of imposing on decision makers different *choice architectures* (Thaler and Sunstein, 2008).

The most widely known context effect is the *Asymmetric Dominance Effect* (ADE), also known as *Attraction* or *Decoy Effect* (Huber, Payne, and Puto, 1982; Huber and Puto, 1983). Consumers who are subject to the ADE are more likely to choose a target product rather than its competitor if the target is presented alongside a decoy product that is clearly dominated by the target, than if it is presented against the competitor only. That is: adding an asymmetrically dominated option, that should be irrelevant to the decision maker, makes it more likely that the now asymmetrically dominant option is chosen, at the expense of the remaining one.

The ADE has been widely replicated in the marketing literature – indeed, it is arguably the biggest export from the marketing literature to the rest of the social sciences and beyond. Alongside dozens of studies in marketing, also biology studies of different species of animals (Schuck-Paim, Pompilio, and Kacelnik, 2004; Shafir, Waite, and Smith, 2002) have replicated it. The literature in economics is scarce, but it does include a paper focusing on lottery choices (Herne, 1999).

The ADE has been and still is of great interest to theorists, be they economists or psychologists, since such a widespread violations of a core axiom is sure to spark interest in theories able to account for it. Finding a robust, replicable exception to standard decision theory was the stated aim of Huber, Payne, and Puto, 1982. In order to achieve this, and according to the best scientific practices, the simplest possible design was adopted. The original experiment (and the dozen of subsequent replications and extensions) adopted a between-subjects design, in which a group would face a set $\{X, Y\}$ and another group would face a set $\{X, X', Y\}$ in which X dominates X' . Differences in the relative choice share of X and Y in the two groups, provided that X and Y had similar choice shares in the population would be interpreted as violations of *regularity*, a large-sample corollary of the IIA axiom. This is the result that was replicated dozens of times in different contexts, over different populations, animal species, and set of stimuli.

The result is interesting for theorists – the violation is robust and pervasive – but, because of the experimental design used and of its limits, it is of little interest to anyone else. First, it is just an aggregate result, and does not shed any light on the personal characteristics correlated to the effect. Second, it relies on a between-subjects design, hence on comparability of preferences

across groups. Due to this constraint, the effect has been studied only in situations in which the decision maker is close to indifference between the alternatives proposed. For an economist, this is disappointing: we go at great lengths to find an effect that matters only in those very circumstances – indifference – in which *any* cue might tip choice one way or the other. Third, and foremost for what concerns my work, the experimental design chosen provides a proof of *existence* of an effect, but cannot provide a *measure* of its intensity. Without a precise, individual-level measure of the ADE, it is difficult to translate the core result to practical marketing or policy applications, or trace its consequences to the supply side, building strategic interaction IO models on it.

The aim of my work in the domain of context effects and the ADE – a work carried out jointly with Alexia Gaudeul – has been to overcome the limits of the traditional design, and to move towards a *measure* of the ADE (and, in principle, of other context effects). These new measures are instrumental to two main overarching goals: enter the debate about consumer sovereignty, freedom of choice, paternalism and nudges on the one hand, assessing if and to what extent people are rational being to which we should delegate all freedom of choice; and to investigate the role of context effects in shaping firm behavior and give insights for competition policy. In the following I will give details about the new measure of ADE we devised and about a first application of it in the realm of behavioral IO and firm competition.

Measuring context effects at the individual level

All happy families are alike; each unhappy family is unhappy in its own way.

Lev Tolstoy

The first step of the research project consisted in developing an experimental design able to yield individual-level estimates of the ADE, measured in a simple and powerful unit: money. The question answered is: how much money do people leave on the table because they are subject to the Asymmetric Dominance Effect?

We needed three separate steps to reach our goal: move from a traditional between- to a within-subjects design; find a way to measure the effect in utility space – operationalized as money; and devise an easy estimation technique that would yield robust and unbiased individual estimates. To reach the first goal we simply gave the subjects several choices instead of one; to reach the second, we adapted concepts from price-list designs, varying the utility of the target and the competitor around the indifference point to measure preference – one way or the other – in a continuous way. The latter point proved trickier, and we had to cross discipline boundaries and look at psychometrics to find a suitable estimation technique that was at the same time easy to implement and would yield individual level estimates.

The ADE is the result of two crucial factors being present: a not trivial *trade-off* between two product characteristics (i.e., location and size for apartments; duration and comfort for train trips; price and quality for food products; and so on...), and the presence of a *decoy*, asymmetrically dominated option along *one* of the two dimensions, making comparisons along the remaining dimension easier. In a way, ADE relies on enhanced comparability: by making two of the three options more comparable, it gives salience to the dominant-and-comparable option,

that is, in a simple heuristic, 'at least better than another option' (Gigerenzer and Goldstein, 1999).

The most crucial aspect of the trade-off aspect is that the utility computation of this trade-off happens, unobserved, in the mind of subjects. Since the utility function of subjects is not observable, it is not possible to devise an elicitation mechanism that could fully reveal the individual ADE to the researcher. To overcome this issue, together with Alexia Gaudeul we devised a task in which the two dimensions are *spurious* – i.e., only *one* dimension is relevant, but it is 'unpacked' into two, and it is not trivial for the subject to invert the function and go back to the unique, underlying dimension that matters. In our experiment, the underlying dimension that matters is unit price; and the two-dimensional space in which it is projected is given by posted price and size (think of weight, or volume for real products). In the simplest possible example, we could ask subjects to choose between one liter of water for 4 Euro and 0.33 liters for 1.

The clear benefit of our shift from two true to one true, two spurious dimensions is that we have an *objective standard* to evaluate subjects' performance. The goal – expenditure minimization – is clear and unambiguous to subjects; behavioral effects like the ADE can only *reduce* utility when compared to maximising behavior. Moreover, once we can measure the monetary (utility) difference between any two options, it is easy to just let it vary around the indifference point in order to estimate a monetary measure of the ADE (and potentially other context effects).

In Crosetto and Gaudeul, 2016 we hence gave each subjects several choices, in an expenditure minimization task. The metaphor used was one of buying paint to cover a given wall. Subjects were faced with a discrete choice among three options, each of them represented by a price and by an area that the paint bought could cover for that price. Areas were shown as regular shapes (circle, square, equilateral triangle) of different sizes. An example task is given in Figure 11.

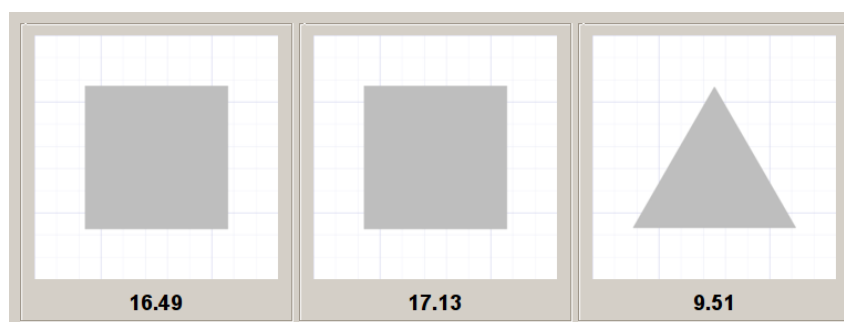


FIGURE 11

To a subject that could perfectly compare shapes and sizes, the experiment consists of a one-dimensional problem only: unit price. To regular subjects, though, the task is way trickier and is fully a two-dimensional task: shape and size on the one side, and posted prices on the other. We varied the underlying unit prices so that the relative unit price of the target – the asymmetrically dominant option – was higher equal, or lower than the one of the competitor – the non-comparable, non-dominant nor dominated option.

If they follow the ADE, subjects disproportionately prefer the target – in Figure 11, the left-hand square. But the traditional ADE design can only show this fact, without measuring it;

and, traditionally, it has shown it only in contexts in which the utility of the target and the competitor (in the Figure, the right-hand triangle) is the same. If the subject still prefers the target when its underlying unit price is higher than that of the competitor, then it means that the ADE is relevant away from indifference; and the amount at which indifference is shown gives us a new, monetary measure of the ADE.

Given this task, how can we come up with the appropriate estimation technique? Thanks to Pete Lunn (Lunn and Somerville, 2015) we discovered the psychometric function (Klein, 2001; Wichmann and Hill, 2001). It is a parametrically estimated sigmoid function that relates the subject response to an independently varying stimulus. In our data, the response is the choice of the target, and the stimulus is its utility premium w.r.t. the competitor. For a perfectly discriminating subject the function crosses the 50% probability of choosing the target at the point of zero utility difference, and it has a steep slope. The flatter the slope, the lower the precision of the decision maker. If asymmetric dominance has an effect, then the function crosses the zero-difference point at a probability higher than 50%.

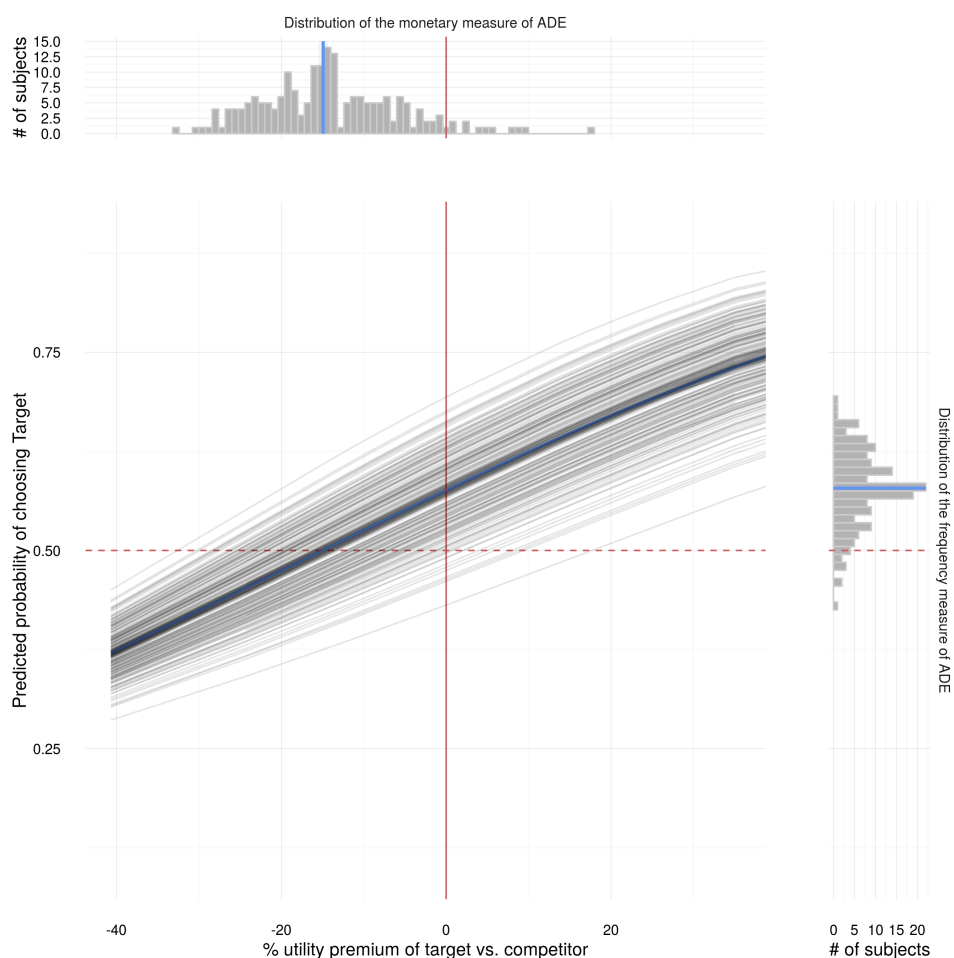


FIGURE 12: Estimated psychometric function from Crosetto and Gaudeul, 2016

The *vertical* distance on the zero-difference line between the estimated function and the 50% probability line is the traditional – *frequency* – measure of the ADE. At the individual level, this

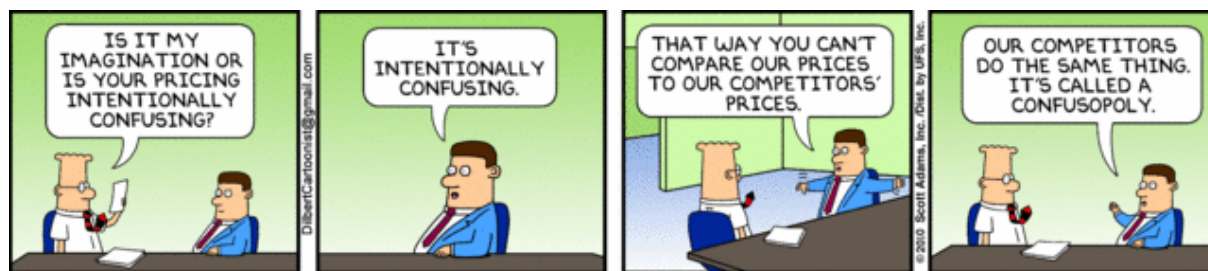
represents the likelihood of choosing the target when it gives the same utility as the competitor. At the aggregate level, this is the share of subjects choosing the target. The *horizontal* distance on the 50%-probability line between the estimated function and the zero-difference line is our *monetary* ADE measure. It represents the money left on the table due to the attraction effect. A positive *monetary* measure means that the subject is leaving money on the table by choosing the asymmetrically dominant option.

This psychometric function can be easily estimated using a mixed-effect logit model. Results of this approach for our experimental sample of 202 subjects over 18 choice tasks is given by Figure 12. On average, our monetary measure shows that subjects still prefer the target up to the point where it is 14% more expensive than the competitor.

Three results are noteworthy. First, subjects *do* leave money on the table because they are affected by ADE; they can indeed be exploited by clever use of context effects by firms, up to a non-negligible point. Second, there is substantial between-subject heterogeneity: some subjects are affected by a reverse-ADE effect as they are estimated to prefer the competitor over the target and to pay a price for this. Third, the ADE problem is truly two-dimensional, since subjects are heterogeneous both in intercept *and* in slope, meaning that subjects showing the same frequency-ADE in conditions of indifference show different monetary-ADE – leave different amount of money on the table.

We are the first to show such a measure, and in the following we will show one use of these results in Industrial Organization, and the first preliminary results of a further paper extending the measure to new stimuli and adding a time dimension.

What is the effect of ADE consumers on firm strategies?



Once having measured the ADE, the next logical step was to track its possible consequences on firm behavior. A growing field of study applies behavioral insights to Industrial Organisation models – to investigate how and in which ways assuming boundedly rational consumers impacts firm strategies, competition and collusion (Ellison, 2006; Kalaycı, 2011; Kalaycı and Potters, 2011; Spiegler, 2006, 2011). When faced with less-than-rational, confused consumers one possible strategy, as in the Dilbert comic above, is to collude in confusing consumers. Another possible strategy is to compete to attract the wiser consumers – and as an unintended consequence, make the market a better place also for naive or confused consumers that would be easily exploited if left alone.

The policy implications of such strategies are relevant. If firms engage in confusopolies, then policy interventions, nudges, mandated standards, are justifiable and welfare-enhancing; if,

on the other hand, a small share of wise consumers is enough to drive firms to compete, then deregulation should be preferred from a welfare point of view.

We set up a theoretical and experimental investigation in this domain in Crosetto and Gaudoul, 2017. We set up a spuriously differentiated economy à la Perloff and Salop, 1985 with switching costs and allow the consumers to be of two kinds: naive, brand-loyal consumers that cannot exploit similarities between products even when they exist; and more sophisticated consumers that do exploit dominance when they see it but are on the other hand subject to the ADE, roughly in the proportions and with the intensity shown in the paper described above.

While the duopoly model is easily solved, the triopoly – the simplest model allowing for the ADE to play a role – is characterised by Edgeworth Price cycles (Doyle, Muehlegger, and Samphantharak, 2010), has no Nash equilibrium in pure strategies and its unique mixed strategy must be computed numerically. We also solve the model assuming reinforcement learning, imitation, myopic best reply, and a form of tacit collusion. We have once a set of stylized facts for the five theoretical approaches, and we set up an experiment to see how subjects behavior compares to these predictions.

Before moving to results, one issue has to be dealt with. Does it make sense to run IO experiments in which students play the role of firms? It might be argued that firm executives face incentives that crowd out the behavioral biases that characterise subjects; and that hence the external validity of IO experiments is limited at best. Nonetheless, it looks as if a certain market moral code crowds out reciprocity and other-regarding preferences when running market experiments (Holt, 1993; Roth et al., 1991). Behavioral IO experiments are now routinely run (Kalaycı and Serra-Garcia, 2015; Shchepetova, 2012), and we decided, despite our doubts, to join in the show.

In our market, firms choose how to present their products in addition to choosing their price. We implement spurious complexity on an otherwise homogeneous good. This allows firms to play the confusopoly strategy if they so wish – by keeping prices high and choosing heterogeneous formats that the consumers fail to compare, they can each secure their share of loyal consumers. Alternatively, they can compete for the few, savvy consumers and compete on prices by setting the same way of presenting their products and hence make apparent that the products are indeed the same. Demand is simulated – subjects only play the role of firms. We study the impact of varying the share of savvy consumers (0, 10% and 20%), and their degree of ADE (0, 10% and 20%), following results from the previous paper.

An old, counterintuitive theoretical result due to Møllgaard and Overgaard, 2001 shows that more transparent markets in which competitions for savvy consumers is potentially tougher are paradoxically the ones in which more collusion is observed. This is due to the fact that when products are easily compared, price wars tend to be very harsh, as consumers switch easily away from the most expensive products. Harsh punishments are one of the conditions for sustained collusion: a deviation from a colluding strategy can be easily sanctioned with a sudden costly war, and the tougher the punishment, the easier it will be to maintain collusion, as in a mutually assured destruction long-run equilibrium.

The behavioral patterns of our experiment are consistent with tacit collusion, and the more so the more savvy consumers exist in the market. To get a feeling for this result, raw data from a session with 20% sophisticated consumers that are subject to a 20% ADE are displayed in

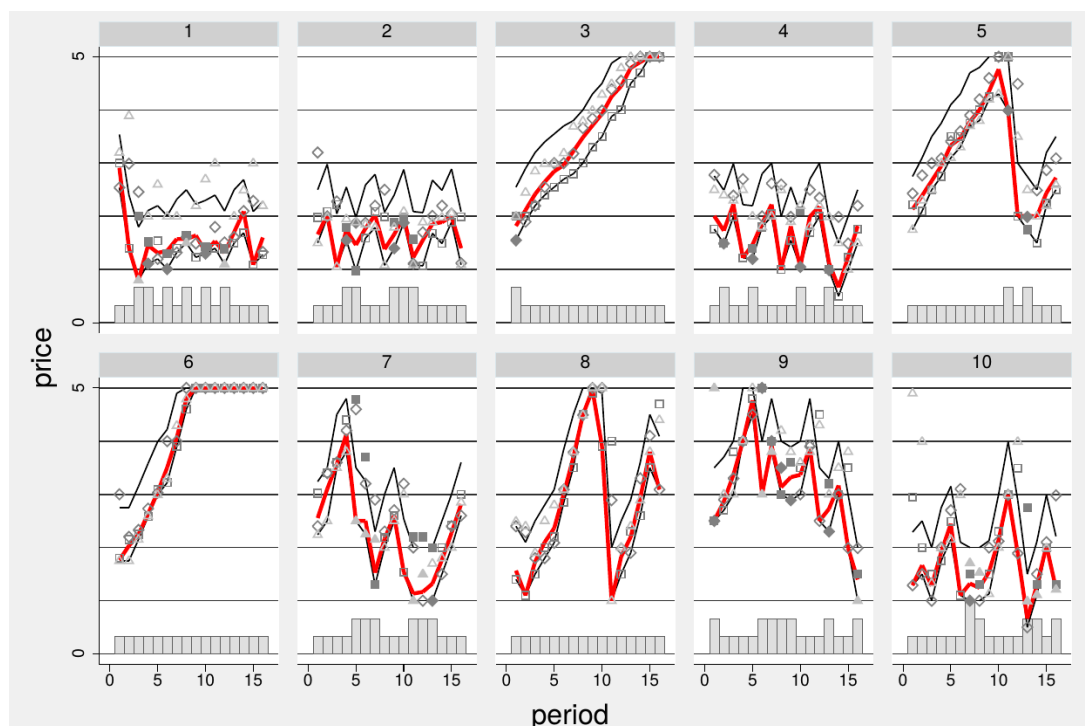


FIGURE 13: Raw data from Crosetto and Gaudeul, 2017 showing evidence of collusive behavior

Figure 13. The red line shows (smoothed) prices; the bar on the bottom of the plots represent the number of firms adopting a common format (by default, one firm always does); and each subplot is a group of three firms – triangle, square and diamond. There is no end effect as the game was indefinitely repeated, with a random stopping point. Groups 3 and 6 clearly show the establishment of collusion, with prices slowly rising to the maximum available. Groups 5 and 8 show a similar pattern, but with a deviation from a firm – in group 5 price cuts coincide with the adoption of a common format for the product – followed by a brief price war and the return to cooperation. Other groups clearly try to collude (7, 9, 10) but fail more often as we see more price wars – and the associated use of common formats.

The main result – there is widespread tacit collusion, and it increases in the share of savvy consumers able to sanction firms that do so – holds only when firms can perfectly observe each other. When they don't, it is harder to detect and punish a deviation, and collusion is less prevalent. Our findings provide a potential explanation for how Coca-Cola and Pepsi are able to keep profits high. Those two firms have been able to maintain a high degree of perceived differentiation and have avoided downward spirals in prices, despite essentially selling the same product: blind-test subjects routinely cannot identify which is which. Both firms avoid competition by focusing on different consumers, sponsoring different sports, differentiating their logos, choosing different colors for their packaging. They devote most of their efforts to maintaining the illusion of a difference between their two products. From this point of view, the soda market looks indeed like a *confusopoly*. Confusopolies might be more widespread than we thought. It pays out to build one, and the presence of savvy consumers makes paradoxically the task easier.

Food choices

I started to do research on food research when joining INRA in September 2013. While it is a field somewhat removed from my former research interests, it has characteristics that make it particularly appealing to an experimental economists interested in individual heterogeneity, measure and methodological innovation. Food choices are everywhere. They are made in differentiated contexts at a very high frequency. Choices respond both to immediate rewards – hunger, pleasure – and long-term goals – health, weight control. Food items have such a long list of intrinsic characteristics, credence attributes, social values, that they are the perfect testbed for many a behavioral theory of decision making. As an added bonus, the agricultural economists involved in food research have long been interested in measuring willingness to pay for several different attributes, using a variety of clever elicitation methods, and hence provided me with an opening on a mature and diversified literature using tools I did not yet know. And finally, research on food, closely connected with policymakers and producers, gives a researcher that very unusual vertigo – the feeling that the research we do has a *real* chance to impact real-world decisions.

One of the key point about the economics of food consumption, is that, theoretically speaking, there is not much of it. A very traditional approach of applying standard consumer theory tools is followed in this field. This means, in applied work, computing huge cross-elasticities table from consumption data, and then use those to simulate policy impacts. This is in stark contrast to the marketing and consumer research literature, in which the particularities of food choices have been widely explored – as is the fashion there, in a mostly a-theoretical setting.

Recently, experimental research has entered the food domain, in a way bridging a gap between economic theories on the one side and marketing and psychology research. Experiments have been so far carried out mainly on three topics: eliciting risk and time preferences of farmers, as a way to assess the likelihood of adoption of new agricultural products and of compliance with policies (Jimenez, 2003; Nielsen, Keil, and Zeller, 2013; Reynaud and Couture, 2012); eliciting the willingness to pay for credence attributes of food – origin, corporate social responsibility, GMO content, social value (Botzen and Bergh, 2012; Drichoutis, Lazaridis, and Nayga, 2009; Drichoutis, Nayga, and Lazaridis, 2012; Lusk, 2003; Lusk and Coble, 2005; Lusk, Norwood, and Pruitt, 2006; Lusk et al., 2006; Roosen, Lusk, and Fox, 2003); and public policy evaluation.

This third stream of research is the main focus of the Grenoble team I joined. In food, public policy mainly revolves around labels (Bialkova and Trijp, 2010; Drichoutis, Nayga, and Lazaridis, 2011; Gorton, 2007; Grunert and Wills, 2007; Vyth et al., 2012). Labels are interesting for behavioral theory and experiments because their efficacy depends on how well subjects interpret information cues and integrate them into their decision processes. When working with food labels, researchers have to integrate in their models beliefs, attention, cognitive limits, and heuristics.

For all its interesting sides, though, the field of experimental food economics research has been using for years a limited set of tools. Becker-DeGroot-Marschack value elicitation and Vickrey N-price auctions are the workhorses used in most papers eliciting willingness to pay for different food attributes, food innovations, and to evaluate the possible impact of labeling policies, regulations, and bans. When dealing with risk and time preferences of farmers, the reference

tools in the experimental economics literature are used, mostly with no or very limited adaptation. In other words, the field of experimental economics applied to food consumption is now mature. The core methods have been fruitfully used for years, to the point that WTP-experiments have now successfully moved beyond academia, and are routinely run across the globe by consulting firms and public bodies.

When moving to Grenoble, I had the chance to join a team that had been innovating on methods for the best part of two decades before my arrival. The key innovations brought about by the Grenoble group – Bernard Ruffieux, Anne Lacroix, Laurent Muller, Stéphane Robin – spanned in two directions: the shift of the focus of experiments from a single product to a whole diet, with the possibility to look at substitution effects and to fully integrate the insight that eating choices are very much alike to portfolio choices; and the introduction of cognitive limits and heuristics in the diet-building exercises. My own contribution has mainly been to help with the development of new, robust elicitation methods, and to integrate the role of social cues and context effects in food choice.

Evaluating the impact of labels on diets: e-shopping experiments



The main public policies priorities when dealing with food choices these days are food safety and health concerns. Rising rates of obese and overweight people cause concerns to the long-term sustainability of the sanitary systems, and for the overall health of citizens in the developed world. The two main policy tools have been regulation of the supply side – e.g., mandating lower amounts of salt, sugar, and fat in recipes – and reduction of the information asymmetry between producers and consumers, by means of compulsory information standards or (usually voluntary) labeling schemes.

The usual approach in evaluating ex-ante the impact of labeling schemes and other demand-side policies is to implement simplified choice environments, in which consumers have to rank two or three products according to their perceived healthiness. A continuous analog of this approach is to elicit the WTP for a product without and with the label, and to draw conclusions about the value of the added information from comparison of the two WTPs. However, building a healthy diet is a different task than facing binary choices. In some sense, eating is akin to portfolio management: mixing in a large portfolio one or two risky assets means taking very little risk indeed; in food, no single item is going to irreparably damage your health, irrespective of its level of fat or salt or sugar.

This is the main reason why the Grenoble experimental economics group has been working on diets – operationalized as daily baskets of products, or choices from a cafeteria-style daily or weekly meal planning – rather than single product choices for the best of the last decade. The main methodological workhorse is a diff-in-diff design, in which subjects are asked to shop for one or more days on an e-shopping environment; then an information or price shock is added – a label (Muller and Ruffieux, 2012), information about the environmental impact of food items (Ruffieux et al., 2014), taxes and subsidies (Darmon et al., 2014; Muller et al., 2017); and finally subjects are asked to shop again. This design makes a clean diff-in-diff, across-treatment comparison possible, exploiting the (within-subjects) changes between phases 2 and 1 and the orthogonal between-subjects treatment manipulation.

I had the chance to participate in two studies applying this general method: Crosetto, Muller, and Ruffieux, 2016b and Crosetto et al., 2017 (these are in French only; English versions are forthcoming).

Both studies share the same methodology. A sample of subjects from the general population is invited to the laboratory for an artefactual field experiment. Subjects are given a show up fee of about 30 Euro and asked to shop, following their usual shopping habit, keeping in a mind a fixed period of time (in these studies, two days). They are exposed to a paper catalog of about 300 products, covering 27 grocery categories and displaying for each product a picture, weight or volume, and price. Prices are the same found in shops in Grenoble, as queried before the sessions run. Subjects can then shop on an e-shop custom online shopping environment, thanks to which they can also access extra information about ingredients and the nutritional table. The experiment is incentive compatible: subjects know that they are submitting binding choices, meaning that they will have to buy from the experimenters the items they select *if they are available*. In a separate room, we stock about one quarter of all the catalog products. The intersection of the items selected by the subjects and what we have in store is then sold, at the catalog prices, to the subjects at the end of the session.

Subjects are asked to shop twice; once on a neutral catalog, displaying products and prices; and a second time after a nutritional label has been added to all the products of the catalog (and on the e-shop interface). One of the two carts is randomly selected at the end of the experiment to be binding. Both Crosetto, Muller, and Ruffieux, 2016b and Crosetto et al., 2017 deal with nutritional labels, and in particular with a new front-of-pack label, named 5C or Nutri-Score, developed by the French Institute for Health and Medical Research INSERM, Ducrot et al., 2015.

A very short note about the policy context is in order. The French Ministry of Health under François Hollande, Marisol Touraine, put among her top priorities a new front-of-pack label that could be more intuitive for consumers. INSERM developed a new, 5-color label called 5C. 5C is based on the UK's Food and Standard Agency Nutri-Score nutritional score (Rayner, Scarborough, and Lobstein, 2009); it is an aggregated label, giving an overall assessment of the product rather than a breakdown of the food nutritional components; and it conveys the information much alike the energy efficiency label used in electric home appliances, as a color on a 5-color, green-to-red scale (Figure 14, center). The Ministry of Health asked us to perform a first experimental test of the label, with the aim of pushing the 5C/Nutri-Score label as the one recommended in the then forthcoming 2016 Health Bill (Loi de Santé). We compared the 5C with a Guideline Daily Amount – GDA, displaying for each of three nutrients, fat, sugar,

and salt, the contents of the food item as percentage points for the recommended daily amount for an average adult, Figure 14, bottom left – and a Traffic Light label – TL, displaying the very same information by means of a traffic light color code, Figure 14, bottom right.

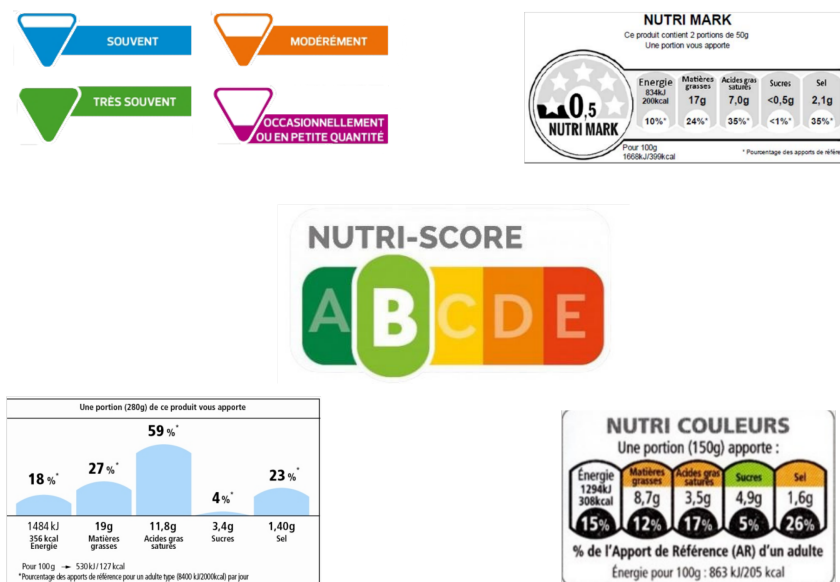


FIGURE 14: Labels tested in Crosetto et al., 2017

Results, detailed in Crosetto, Muller, and Ruffieux, 2016b, showed that all three systems improved the nutritional quality of the average subject's shopping cart. 5C and Traffic Lights resulted in a significantly greater nutritional impact than the one obtained imposing a GDA label.

The ministry failed to push the 5C label as the official French recommendation, as a consequence of an intense fight by firms – both producers and retailers – that stood against a color-code label claiming it misrepresented perfectly fine foods as traditional cheeses and cured meats as bad for people's health. The law, instead, mandated a large-scale randomized control trial (RCT) to be run in order to assess the most performing out of a battery of labels. This RCT, involving 60 supermarkets in two regions ran in December 2016, collecting data from about 150 thousand fidelity card owners of the supermarkets involved. Given the enormous cost of such a study, labels were applied on 4 shelves only among the dozens present in the supermarkets, and only on those products for which an authorization by the producer had been granted. Parallel to the RCT, our group in Grenoble was involved to run a lab experiment following as closely as possible the methodology of the RCT, but also using our diff-in-diff e-shop design. We hence ran the biggest experiment ever run in our laboratory, involving 809 subjects over 7 treatments. We pitched one against the others the 5 labels of Figure 14, plus a neutral benchmark treatment in which subjects had to shop twice without labels, and a Nutri-Score treatment in which we labeled in phase 2 only the (very limited) subset of products that happened to be labeled in the RCT. The newly-introduced labels were the SENS (Figure 14, top left), developed by the firms, and based on frequency recommendations rather than judgments of product healthiness; and Nutri-Mark (Figure 14, top right), used in Australia and New Zealand.

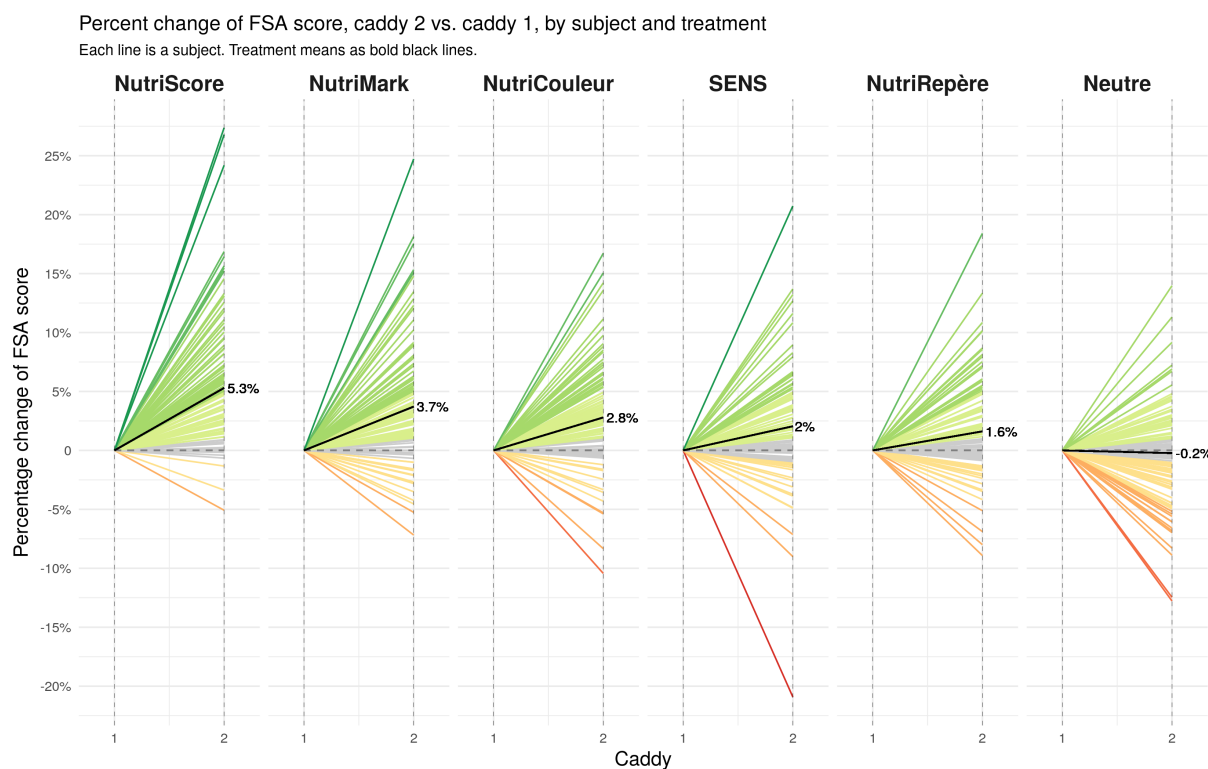


FIGURE 15: Percentage FSA score nutritional improvement, Crosetto et al., 2017

Results, detailed in Crosetto et al., 2017, showed once more that the NutriScore label outperformed other labels. An outlook of the results is given by Figure 15. The figure depicts percentage changes in the nutritional quality of the shopping cart of each subject from the first phase of shopping (normalized to zero) to the second phase, by treatment. Each line is a subject; improvement in nutritional quality are given in green / light green and reductions in orange / red. Nutri-Score (significantly) outperforms all other labels, followed by the other aggregate label, NutriMark; all others follow, with the neutral benchmark treatment well doing its job and showing disperse, but nutritionally neutral behavior in cart 2 compared to cart 1. The results of the RCT were very virtually identical in terms of ranking, but much smaller in terms of effect size.

As a result of the coherent and strong results of the RCT and of our lab experiment, the French Ministry of Health adopted the NutriScore as the official recommended label in France; the label has already been adopted by some retail giants as Auchan and Leclerc, among others.

All in all these experiments have taught me that ex-ante public policy evaluation in the lab is possible. The efforts required are high – we had to track, buy, stock and sell *hundreds* of products in each session – but the results are encouraging. Moreover, the close proximity of our last experiment with the publicly-funded RCT allows us to make what I think is a major methodological contribution to the field of experimental policy evaluation: we are writing a paper tracking down in detail similarities and differences between a huge field RCT that cost about 4 million Euro and a large, but still limited, laboratory experiment that cost about two orders of magnitude less. Results are similar in nature, sign and ranking; dissimilar in magnitude. The

work is still ongoing on this methodological comparison, but the stage is set for its main result: artefactual field experiments seem to be a reliable low-cost alternative to RCTs, yielding greater control and power at much lower cost.

Assessing the role of cognitive limits in label use



The second front of my research on food has dealt with integrating the role of cognitive limits into economics food research. Food labels can be seen under (at least) two competing perspectives. When seen from a *homo oeconomicus* perspective, labels are a means to convey information and hence to at least partially correct the deep information asymmetry present in food choices. If we assume a decision maker to have fixed, stable and complete preferences, then giving more information is by definition welfare enhancing, since a subject can only (weakly) benefit from having more elements upon which to base his (rational) choice. When seen from a *bounded rationality* perspective, though, labels are a means to convey a message that might impact the decision maker over and beyond the simple information they contain. Colors might induce halo effects and trigger heuristics; simple messages (organic, local...) might simplify search or *nudge* consumers in a predetermined direction.

Identification of the *best* label depends on the assumptions we are willing to make on consumers and firms. If we think that consumers tend to be rational and make good use of all information available, then it makes sense to give them as much information as possible, and leave to them to parse it and integrate it in their choice processes. A detailed nutrition table would be best. If instead we think that consumers are subject to biases, follow fast and frugal heuristics and tend to be overwhelmed by lots of information, then it makes sense to propose aggregated labels summing up in a salient design all relevant information. Of course the second approach lends itself to paternalistic practices, since the decision of the direction in which to nudge consumers is taken by third parties (or, in the case of marketing nudges, by the firms themselves). There is a trade-off between the granularity and level of detail of the information, on the one hand, and the simplicity and intuitiveness of the message displayed, on the other. In order to identify the best label, we have both to make assumptions about consumer behavior and to test those assumptions in an environment as similar as possible to the one faced by consumers when choosing which food to buy for their diets. Similar work is carried out in Grenoble about the reaction of firms to labels (Bonroy and Constantatos, 2015) and in principle IO experiments in this domain are also possible.

The experimental literature on labels (see, among others, Borgmeier and Westenhoefer, 2009; Grunert and Wills, 2007; Hawley et al., 2013; Kelly et al., 2009; Moeser et al., 2010; Synovate, 2005; Vyth et al., 2010, 2012) is vast and valuable, but suffers from two key methodological issues. First, evaluations of performance are based on products rather than diet. While the aim of the labels is to help consumers build better overall diets, the proxy used by researchers is to ask subjects to rank two or three products. Second, the question asked determines the relative performance of the labeling schemes. Ordinal questions and relative rankings favor color-code labels, since those give away the answer cleanly and without any clutter; cardinal questions and absolute rankings favor numeric labels, whose discrete nature provides easier guidance.

But how do labels perform when asking a more directly relevant question, i.e. to compose a diet – a basket of differentiated food products consumed over the course of one or more days? In Crosetto, Muller, and Ruffieux, 2016a we built a new experimental design that allows us to a. assume away individual preferences; b. directly observe diet-building; and c. involve hundreds rather than isolated pairs or triads of products.

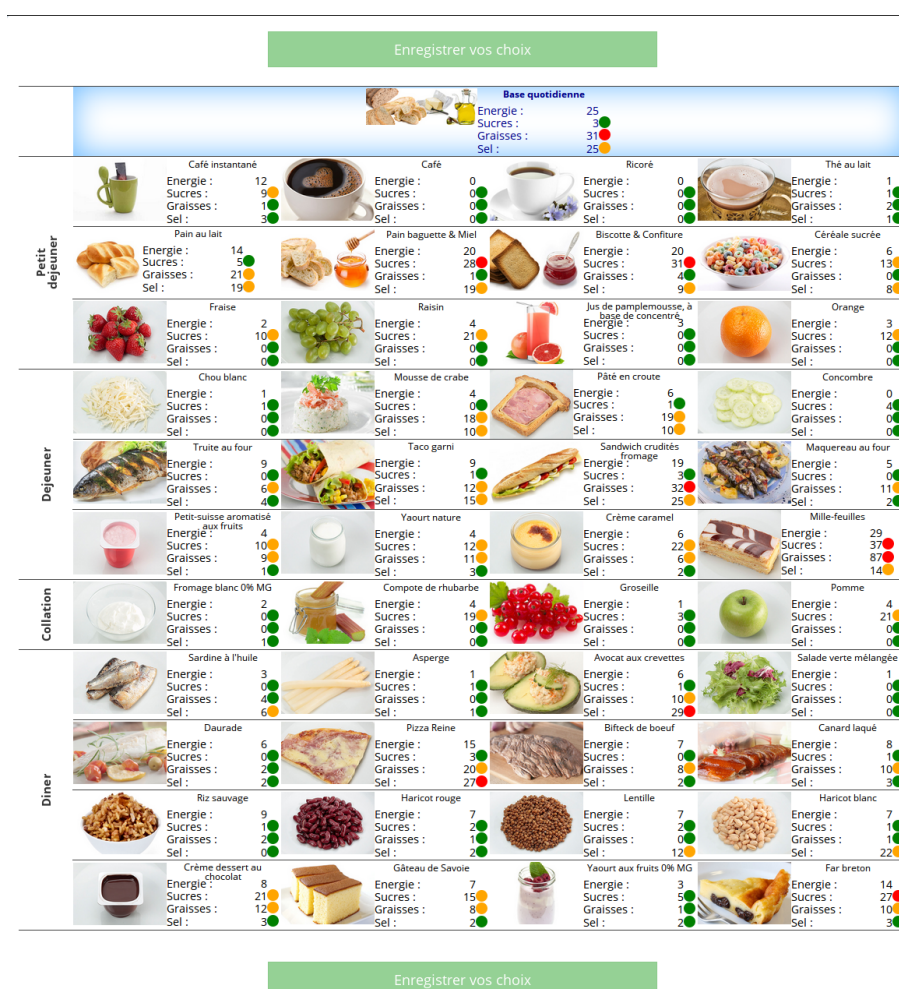


FIGURE 16: Diet-building task from Crosetto, Muller, and Ruffieux, 2016a

In the experiment, we ask subjects to act as the hired nutritionist of a canteen that caters for people for the whole day – breakfast, lunch, afternoon snack, dinner. Subjects must compose

daily menus, and face 4 cafeteria-style ready-made cooked options for each of twelve courses during the day (Figure 16 shows a screenshot of a task). Subjects must not choose their preferred meals, but rather daily menus that comply with one (Kcal), four (Kcal, sugar, salt, fat), or seven (all of the above plus vitamin C, calcium and fiber) constraints. If the menu they create satisfies the constraint, they are paid a fixed fee; otherwise, zero. The exercise in all respect identical to a linear program, as the aim is to maximize under constraints. Preferences should play no role. Given this baseline scenario, we implement treatments giving the subject no information (just the product images, to control for beliefs); or giving them GDA-style (numbers), TL-style (colors) or combined (number+colors) information. This design allows us to show the *pure* effect of labels once preferences are irrelevant and in a controlled setting. We recruit students as well as general population samples, and in a separate experiment we limit the time allotted to each task to 2 minutes – about 70% the median time of the non-timed sessions – to induce some cognitive pressure on the subjects.

Results show that numeric labels like GDA perform better than color-codes like TL when subjects do not face time constraints. This is especially true for highly skilled, math-oriented engineering students, and less so for the general population. When time is limited however, TL and GDA have identical efficacy with 4 nutritional goals, and TL even outperforms GDA with 7 nutritional goals. Analytical and numeric label impose too much of a burden on subjects as the time allotted to each decision shortens and the number of (possibly conflicting) goals increases. Given our artificial setting, and the fact that our artificial goals were set in terms of GDA, our experiment is a very favorable terrain for computations and, therefore, arguably best suited for GDA. If GDA does not outperform simple color codes in our biased environment, with incentives and focalised attention, it is likely to fail in the noisier and fuzzier real world.

Where do we go from here?

Cognitive limits, context effects and the challenge of convergence

I also would never refer to anomalies; what the notion of an anomaly says is that we have the norm of rational behaviour and you somewhat deviate from this norm. But actual behaviour is not just one step away from this kind of norm - it's a universe of steps away.

Werner Güth

Possibly the most discouraging finding from my work on risk is that as a discipline, we experimental economists seem to have reached a blocking point on our measures of risk. Different tools yield estimates that correlate poorly and sometimes flirt with orthogonality. Risk preferences seem to be volatile, inconsistent within and across tasks. Paper upon paper trying to compare measures finds results that are hard to rationalize into a common framework. Time preferences – a field in which I have not worked but that I know from reading and attending conference presentations – have similar issues. When trying to apply risk and time preferences to food issues, the same frustration of fragile results hangs on the literature: elicited preferences correlate poorly, if at all, with food and health behavior. Our basic tools as economists - preferences, utility, risk aversion, time discounting - even as refined by experimental and behavioral work, do not seem to be able to predict well eating behavior over neither the short nor the long run. When looking at context effects, the inadequacy of some of the theoretical frameworks and empirical tools used jumps to the eye. Theories of context effects generate mostly from cognitive psychology, and economists have only partly been interested in the topic. For instance, our new ADE measure paper got rejected from an economics journal because, among other things, *'this is marketing, not economics'* and it is one of only three or four papers in total on a very relevant topic for microeconomists.

As a result of these summed frustrations, I have a feeling that we must be doing something wrong. There might be two reasons why measuring risk in a robust way is so difficult, context effects are elusive, and our predictive power in food and other applied behavior is so small. Either our tools are not up to the task, or the theories we use are off the mark. This does not mean that the tools are useless – we learned a lot using them – nor that the theories are wrong; it means that we might not be taking into account important dimensions of the problems at hand. So theories are unfocused, and measuring tools unreliable.

To achieve more focus, a recalibration of existing tools and theories rather than a new foundation might be sufficient. In particular, as a result of my work I think that two key issues should be brought to center stage: a more focused attention to the actual cognitive processes underlying decisions – cognitive limits, limited attention, heuristics – and context effects and

environmental cues. In part, this has already been done. In economics these issues are present since a long time, and have given rise to whole subfields – bounded rationality, agent-based computational economics, behavioral economics, neuroeconomics. A large number of findings from cognitive and social psychology, as well as from marketing and neuroscience, are being integrated in the traditional paradigm, both enlarging and changing it. Yet, most of the tools of the trade are direct emanation of a full-fledged rational choice approach, having been created starting from that benchmark. It is the case of virtually all tools used to elicit value, like the BDM mechanism or N-price auctions; of the designs applied to social dilemmas, as the prisoner dilemma, public good or common pool resource games; of the games used to identify other-regarding preferences, as the trust, ultimatum, dictator games; and of the tools used to elicit risk preferences, as lotteries and multiple price lists.

The workflow followed is rather simple in its weirdness. Experimental economists use these rational-choice, expected-utility maximizer calibrated tools, and then interpret differences in behavior with respect to the benchmark in various ways, towards their goal of pinning down behavioral insights in economic situations. So *overbidding* is the deviation from Nash play in auctions, inequality aversion (or altruism, or reciprocity, or...) could explain deviation from subgame-perfect Nash in the Ultimatum game, and similar considerations plus institutional arrangements help guide the researcher through public good games. This is what Werner Güth, 1995 usually calls *the neoclassical repair shop*. The tools we use force us to see the world from the perspective we are trying to leave behind.

What I would like to do in the coming years is to keep actual behavior *directly* into focus rather than look at it obliquely, as a deviation from a higher form of expected behavior. I think that as a discipline we need to import more ideas from cognitive sciences and to apply them more directly. I think we need methodological tools built no more on the assumption of a unique, monolithic, rational decision maker; but rather on the assumption that people use a variety of heuristics and contextual cues when faced with a problem. And we need to start treating some phenomena as the rule, rather than, as we did so far, as *anomalies*, exceptions to our standard benchmarks. If subjects are never consistent across risk elicitation tasks, it might be because they are looking at the tasks from a very different perspective than the one according to which we created them. If subjects follow information fuzzily gathered from others, even when this information is just a cloud of clicks, then imitation and conformism must be crucial characteristics of our species. If context effects like the ADE are found in several other *animal species*, even some lacking a brain, then this might mean that this type of behavior is built-in deep in our biology.

Given all these findings from mine and other people's work, it does sound rather absurd to keep thinking of them as a *bias*, *anomalies*, just because they happen not to fit with our theoretical assumptions. Rather than being a deviation from the rational paradigm, if taken at face value and looked at with the appropriate tools, these 'biases' could hint at who we really are, how we process information, how we see, compare and choose things. We might want to build new theoretical and empirical tools to take this into account *from the very beginning*, tools built on these insights. Such tools might show us the underlying consistency in behavior that we keep missing – for instance, in the risk domain. Such tools might also be the starting point from which to build a new, more complex but ultimately better theory of economic behavior.

Moving to this new perspective is not an easy task, nor I am saying that my research project for

the HDR is to build a new behavioral theory of the consumer. This is a daunting task that will require the joint efforts of several minds, and maybe a genius or two. But we can get there in steps, and the way has been already paved by several other researchers. Moreover, having an ambitious, long-term, overarching goal that might inform my future efforts seems like a good way to involve motivated students along the way.

And I already started: using new tools, asking new questions that go outside the realm of traditional economic thinking, including other variables. Two projects in particular are at an advanced state. During my last days in Jena I got involved into a neuro-economics projects that is still under way; more recently here in Grenoble I have been trying to look at the effect of social norms and public actions in food choices in a lab context. Even if these projects are not fully in line with the idea of new behavioral tools that could change the way we look at phenomena, they are in that direction, and I will start the discussion of my future plans from them.

To be honest, a drive towards new tools is not only motivated by the big picture of a future, better economic theory, but also out of a practical concern: trying to make my different research interests converge. As you have seen, my research in the last years has touched upon several different themes. This dispersion has been great fun and kept my multi-faceted interest alive. Nonetheless, it is not efficient. As a consequence, the first key task I did set myself on the road to new, *natively behavioral* tools and theories, was to think of an experimental tool that would go in the right direction and with the potential of being applied to several different topics.

As a first step in this direction, I submitted in September 2016 a project for the INRA Young Researcher grant (yes, I am still a young researcher!). The project aimed at developing a new elicitation mechanism based on the fast & slow metaphor and intuition, that would yield both an immediate, intuitive reaction and a colder and slower, deliberate decision on any discrete choice task. Applications to three domains – food choices, context effects, risk attitudes – were planned. The project got approved, and after discussing the two ongoing neuro- and social-papers, I will move to that.

Cautious incursions in the land of neuroeconomists

This house believes that neuroeconomics is not.

John Hey introducing the Oxford-style debate at ESA 2007, Rome

One road that several economists are exploring to overcome the limits of the traditional approaches is to look directly into the brains of subjects Camerer, Loewenstein, and Prelec, 2005; Glimcher and Fehr, 2013; Rustichini, 2015. In a way, this responds to the concerns I raised above about having a theory that is actually based on *who we are* rather than on *who we should be if we were rational*. By directly looking into the brain, we can gain insights about cognitive processes quite literally. The efficacy of the method is disputed, though: t many, it is not clear what can be gained in looking at neural correlates of behavior. To put it like Glenn Harrison did in the 2007 Oxford-style debate at ESA in Rome, *if you want to know where a car is going to, why don't you just ask the driver?* (Harrison, 2008). Besides these methodological debates, one of the main practical problems with this field is that economists are not (or not all of them)

neuroscientists; and that you need a high degree of specialization to understand, conceive and run neuroscience experiments.

While still carrying around all these doubts, I had the chance to participate in a neuroeconomics experiment, in a rather ambitious joint work with Werner Güth that we carried out in Jena in cooperation with neuroscientists, and which is still work in progress for several reasons – not least, the fact that inter-disciplinary cooperation is not an easy task.

In this project we record proposer behavior in a traditional economics laboratory (in Jena), and responder behavior using fMRI brain scans. The focus of the paper was, in line with other works on the motives of sanctioning and the role of voice, on the *impunity game* (Bolton and Zwick, 1995). An impunity game is an ultimatum game in which the responder has no real power, but only voice. That is, she can reject the offered amount, her reject is communicated to the proposer, but this has no monetary consequences for the proposer. Nonetheless, voice comes at a cost, since the responder's rejection does indeed destroy her part of the pie. The impunity game allows us to avoid the ambiguity about the motives of rejections in ultimatum and dictator experiments (see Güth and Kocher, 2014 for a recent review of such experiments). In a dictator experiment the respondent has neither material punishment power nor choice and thus also no voice. In an ultimatum game voice and material punishment go hand in hand. Thus behavioral differences between the two games can be due to any of "no material punishment", "no choice" or "no voice".

We developed an innovative protocol that allowed us to run part of the experiment in a traditional laboratory, part of it in the scanner, and create real choice situations for the subject in the scanner without relying on deception. First, 240 proposer participants of traditional computer laboratory sessions were endowed with either a small or a large monetary pie, and could offer to responders a more or less unfair amount. Then, 24 responder participants took part in a fMRI brain-scan experiment. Each confronted all 240 offers, half of which coming from a large and half from a small endowment, and had to either accept or reject them. Decisions were later communicated to proposer participants via e-mail, ensuring no deception – a feature usually largely overlooked by neuroscientists and neuroeconomists but dear to experimental economists.

Impunity as expected triggers many minimal offers which proposer participants partly expected not to be collected. The overwhelming response was to accept, except for the lowest possible offer which was frequently rejected and more often so when the pie size was large. While brain scan results are still preliminary and depend on our neuroscientist colleagues to be interpreted, the availability of 240 choices for each responder allows us to provide a finer map of respondent behavior than previously done. A minority of subjects is opportunistic, accepting any offer. Most subjects are instead fairness minded, and nearly unconditionally refuse the lowest offers. Moreover, we measure reaction times down to the millisecond. We find, contrary to our expectations, that refusing an offer – an emotional decision – needs significantly more time across all conditions than accepting it – the rational choice in this case. It looks as if time is needed not only to move from an emotional to a more deliberate response; but in settings such as ours, in which accepting unfair offers is the rational thing to do, time is needed to go against one's best material interest to voice one's anger.

A working paper of this project, now long overdue, will be appearing soon. Time will tell if this will be my only – brief, tentative so far not extremely successful – incursion in the land of

the neuro-economists, or if some other things might come out of it.

Introducing social pressure in food WTP

I'll have what she's having.

Old woman at the restaurant in "When Harry met Sally..."

Another line of research that tries to give more context and observe choices in a more realistic scenario is linked to the project FoodPol promoted by INRA in recent years. Within this project, that aimed at studying food policy interventions and their impacts keeping in mind behavioral responses, together with Laurent Muller in Grenoble we focused on the social multiplier effects of labeling. To do so we designed an experiment sitting mid way between a classic WTP elicitation task and an information cascade / bandwagon experiment.

Labeling and food policies do not happen in a vacuum, but are implemented in a social context. Society and culture define what is edible, what is appropriate, what is available and what is perceived as tasty. Moreover, peer effects play an important role in shaping food choices. Consumers might make different choices if they eat on their own or in a social context, or if their choices are or not observed by their peers. Irrespective of the mechanism followed to impact choices, a literature on peer effect in food choice has recently emerged. For instance, social image concerns have been shown to influence the WTP for fair trade chocolate (Teyssier, 2012). Moreover, consumers' WTP might be influenced by the will to conform to a social norm and/or to imitate other members of a group (Arce Salazar and Oerlemans, 2016). Hence, the effect of a label can be magnified or muted by the social context, depending on possible cascade and bandwagon effects. These effects are unlikely to be picked up by the existing WTP experiments, that usually elicit individual changes in WTP following a policy intervention.

Together with Laurent Muller we developed a design allowing us to assess if and to what extent labeling policies have a measurable differentiated impact if experiments are run in isolation – as usually done – or rather in a social context, i.e. one in which WTP choices are observable by other participants. We expose subjects to a product and elicit their WTP (using BDM) first in isolation, and then allowing the subjects to observe the (real) decision at $t - 1$ of a randomly created group. We let subjects repeat the task for 4 periods, in order to leave room for a social norm to emerge. We then introduce a policy – in the form of an information shock – and repeat the procedure once more. This design allows us to observe how the policy affects WTP in a rich way. As in standard WTP experiments, we observe level shifts in WTP. We can also measure the convergence of the WTP within the group, both before and after the shock. The slope of this convergence gives us the speed with which a consensus is formed within the group. This speed can be different before and after the introduction of the policy. We can hence measure the degree of conformism of each group – the degree of convergence to a common WTP within the group, as well as how this degree of conformism interacts with the policy to generate the final WTP.

We apply this method to a dark chocolate bar, a largely consumed product that is usually liked by most people, and to a bottle of wine, a product with a particular social status in France and

that has a high exchange value. We chose an organic label as the information shock, and we incentivize the experiment by means of real sales of both the chocolate bar and the wine bottle.

Preliminary results show that conformism is rather strong, and stronger before than after the exposure to the information shock, showing that consumers *a.* rely on the peers mainly when they lack relevant information about the product, and *b.* have strong priors about the premium they give to organic food, and tend to stick with them. The impact of positive information about the organic nature of the product is, in line with the literature, positive.

A working paper is coming up soon; but over and above what we can report on this paper, the role of social interactions in food choice is a large avenue of research whose feasibility just got boosted by the diffusion of social networks in the population. It is possible to conceive and run ambitious field experiments at little cost, and exploit the viral dynamics of social networks to test the differential effects of policies in a vacuum or in a context with a high level of interactions.

Eliciting fast & slow responses

The division of labor between System 1 and System 2 is highly efficient: it minimizes effort and optimizes performance (...) Laziness is built deep into our nature

Daniel Kahneman

In a recent book, Daniel Kahneman, 2011 puts forth a metaphor, supported by decades of studies, stating that there exist not one, but (at least) two decision systems. System One is fast, intuitive, relies on first impressions and heuristics, and is our default in most occasions. System Two is slow, deliberate, calculating, and enters the picture when called for. Economics has long dealt with System Two only, usually by directly assuming pure System-Two decision makers. Cognitive psychology has long dealt mainly with System One. Kahneman urges us at thinking at them both, and at the way they interact.

The two systems have already been extensively studied, both by psychology *and* by economics. In economics, the usual way to elicit System One responses is by imposing strict time limits on subjects (see, among many others, Kocher, Pahlke, and Trautmann, 2013; Pettibone, 2012; Reutskaja et al., 2011). This way, it is argued, subjects have no time to think and we can assume that they use a fast heuristic. This suffers from a crucial problem: the time limit is exogenously imposed on heterogeneous subjects. For some, it might be too strict – so that only noise and not even a fast heuristic is measured. For others, it might be too long – so that they have time to move to rational behavior.

The aim of my INRA Young Researcher project is to build, test and apply to different domains a new elicitation tool built not more on the assumption of a unique, monolithic, rational decision maker; but rather on the assumption that decision makers use a variety of both fast (heuristic, intuition, emotions) and slow (cost-benefit, deductive) reasoning strategies when faced with a problem. The tool, inspired by an incentive mechanism developed by Caplin, Dean, and Martin, 2011 in the context of search theory and behavior, allows the researcher to elicit both fast and slow responses in a variety of multiple-choice frameworks, and over different application domains. Caplin and colleagues had to incentivize search, and came up with the idea of paying,

rather than the final choice result of a long search process, a randomly drawn choice of the (possibly long) list of provisional choices submitted by a subject while searching. I took their idea, refined it, applied to a multiple choice context, cut its rough edges and made it into what I think is a usable, easy to understand, reliable tool allowing to elicit both fast and slow responses from a subject.

A decision maker (DM) i faces a choice among N alternatives, and is allotted T seconds to make up her mind. Each second t , her (provisional) choice c_{it} is recorded. At the end of the allotted time, the data obtained from the subject is a vector containing all the provisional choices, $C_i = \{c_{it} | t = 1 \dots T\}$. One time point is then uniformly drawn $\bar{t} = \sim U(1, T)$, and the provisional choice recorded at that time $c_{i\bar{t}}$ is binding and determines the DM's payoff. If no choice had been submitted at time $c_{i\bar{t}} = NA$, then the DM is assigned a uniform random choice in the alternative space, $c_{i\bar{t}} = \sim U(1, N)$.

This mechanism sits at the crossroads of traditional, rational choice and native behavioral tools. Its incentive scheme (that assumes self-interested individuals, at least with respect to the outcome of the choice) will induce the DM to use both systems in the task. Since for all the time she spends making up her mind the DM faces a randomly allocated outcome, her immediate interest is to provide a fast, System 1 response that could improve on randomness. Next, since the clock is ticking and each second has positive probability of being binding, the DM must – still under time pressure – consider if its initial choice was the best she could do, and use higher-order cognitive resources typical of System 2 to improve on her choice.

The DM is hence incentivized to give *both* a fast *and* a slow response to the choice situation. The method allows the researcher to reliably observe in one and the same task the use of heuristics, the shift to cost-benefit considerations (if any), and the transition between the two (if any). Of course the DM could behave differently than hypothesized here: he could act randomly all the way through the task, or take his time and submit a *slow* response only. But in any cases, we *incentivize* him to show us a choice *process* and not just a *final* choice. In a way, this design is akin to mouselab or eye-tracking designs. These designs follow what people look at (or click on) as a proxy of the DM's inner decision mechanism. The method proposed here uses incentives to induce the subject to reveal his provisional choices directly, without passing through his eyes or mouse, but using incentives.

The method implements a multiple choice environment. It is highly portable across domains – as all it requires are a set of choices, without putting restrictions on what the alternatives can be (lotteries, products, labels, answers to a quiz, shapes, ...). It can be applied in particular to domains that feature a clear fast-slow cleavage in them. It has both an empirical interest – we can elicit different types of responses without imposing exogenous time limits – and a theoretical one. Most of economic theory concerns itself with *slow*, rational reasoning. Several approaches have sprung up along the decades to model bounded, limited rationality; conditional on his tool giving us starkly different results at different time frames, we might find a way for the two approaches to coexist.

In my quest for convergence, I thought of applying it to food choices, context effects and risk preferences. In the following, I will give details of what has been done and what perspectives are still open.

Food labeling

Food choice is one of the main terrains in which System One and System Two considerations clash. It is eminently a System One affair: most of the time we choose intuitively, we rely on habits shaped by social norms and culture, we are subject to hundreds of stimuli in a few seconds and we usually choose when hungry (Ashton, 2015) – i.e., in a state of relative cognitive impairment. At the same time, food policies, guidelines and even some food labels convey information that would be most relevant to System Two: long-term health-related information that needs to be understood, processed, integrated in our decision patterns, information on location, ingredients, and more. As we have seen when discussing the work we did in Grenoble for the French Health Ministry, this is one of the fundamental motivations of moving from ‘nutrition fact’ tables on the back of pack – that speak only to System Two – to front-of-pack, color-coded information – that has the potential to appeal to System One.

In the literature it has long been argued that color-coded labels are ‘more intuitive’; the other side of the debate has repeatedly stressed that this comes at a cost, as simple labels offers coarse information and are strictly inferior – from a System Two perspective – to the nutritional tables or the analytical labels displaying lots of information. In Crosetto, Muller, and Ruffieux, 2016a we already set up a design to investigate the differential effects of labels and their interplay with cognitive limits. Using the F&S mechanism in this domain can help pin down *how much* more intuitive a label is, with respect to another.

Together with Laurent Muller we ran a first, calibration experiment and will run more in the near future. As done in earlier work, we get rid of preferences by setting up a quiz-like experiment, in which subjects are rewarded for *correct* responses on a specific metric rather than incentivized with the products. The aim was twofold: see the mechanism in action in order to assess if it is well understood by subjects, and give some insights as of the cognitive processing of color-coded *vs.* numeric labels. Coarse labels give less information and might lead to (quick) error; but the converse – i.e. that analytical labels that give lots of information lead to the optimal response – need not necessarily be true. By varying the difficulty of the task and the degree of informativeness and of discriminatory power of the label, we assess if analytical labels actually help, in which time frame, and if a quick-and-dirty first approximation is not overall better than the pondered, costly System Two reconsideration that might (or might not) happen next.

Our sample included 193 subjects, recruited from the general-population. Preliminary results of a label-only treatment – subjects have to compare label without any product attached, the cleanest possible test of the cognitive processes at play – are highly encouraging. Figure 17 summarises the findings of this treatment.

The Figure shows the rate of correct choices in time, over the thirty allotted seconds, by treatment. We ask an *aggregate* question; that is, the question revolves around the general nutritional quality of a product. The aggregate color-coded label (‘o’ in the figure) gives directly the correct solution, and not surprisingly it is the fastest; in this condition, subjects arrive at good rates of correct answers in a very fast span of time – less than 5 seconds. At the other end of the spectrum, the numeric-only label (‘xxx’ in the plot) has opposite properties: it takes a long time for it to be effective – as it mainly appeals to System Two – but it eventually picks up. Interestingly, the mixed label, that combines colors *and* numbers, (‘ox’ in the plot) gives the convex hull of

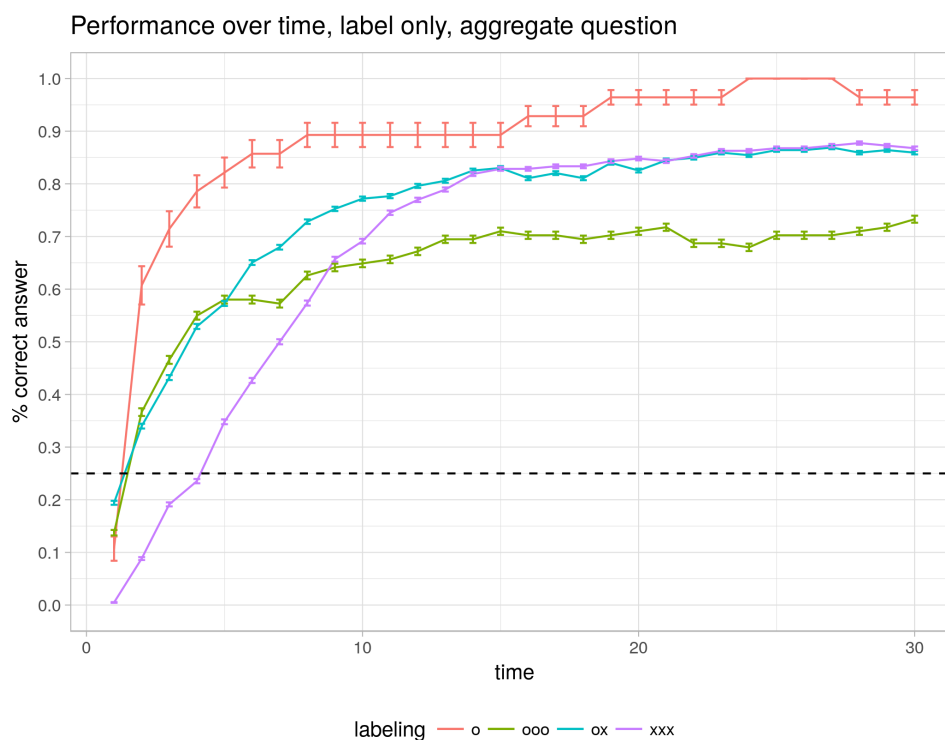


FIGURE 17

the color-only 'ooo' and number-only 'xxx', giving direct evidence that subjects *first* process the coarse color information for a quick-and-dirty approximation, and *then* move to looking at the numbers.

The project just got started and more has to be done before we can reach any conclusion; but it looks as if the method developed can help in 'opening up the mind of subjects', forcing them to reveal at least partly their provisional choices to the researcher.

In the future, I plan to pursue two directions. First, we need to reintroduce preferences. The task described here assumed them away for the sake of clarity. But it is likely that the interaction of cognitive processes and preferences shapes choices. So a first, not immediately intuitive task is to find a way to elicit fast and slow responses in actual choice involving preferences, and still get some clear messages out of the data. Second, we need to find a way to include the role of beliefs and of habit formation. Food choice is fundamentally shaped by the habits that people develop over the years – so much so that I find myself importing Parmigiano wherever I work, irrespective of latitude and longitude. Beliefs are crucial to sort through environments with hundreds of options and abundant information asymmetries. How habits get formed, how they develop, and how they feed in our fast (and slow) choices should be a key point for the economics of food consumption in the near future.

These insights were the reason why, together with Sabrina Teyssier, we organized in January 2017 a Winter Workshop in the experimental economics of food behavior. We think that the inclusion of these issues and methodological evolutions are the key to the future of the field. The Workshop was a success and it will be organized again in 2019; it proved us that there is

interest in changing the rules of the game in economics food research. In Grenoble I will be in a good position to do so, and so will future students.

Context effects

One of the main classes of biases of System One identified by Tversky, 1972; Tversky and Simonson, 1993 is context effects: choice is affected by the addition of irrelevant alternatives to the choice set. Recently the robustness of context effects finding, and notably the Asymmetric Dominance Effect, has been put into question (Frederick, Lee, and Baskin, 2014; Huber, Payne, and Puto, 2014; Simonson, 2014; Yang and Lynn, 2014). In particular, one key issue is to know whether the ADE is a System One or a System Two phenomenon, and if it survives when subjects are given enough time to ponder their decision (Pettibone, 2012). This is of both practical and theoretical relevance. If the effect dies away upon reflection, we could assume its practical relevance in the real world to be limited.

With Alexia we planned to apply the Fast&Slow elicitation technique to our ADE measure. While we were at it, we decided to run a larger study to correct two of the fundamental shortcomings of our previous paper. The most obvious shortcoming was that the stimuli were in reality three-dimensional (shape, size, and price) rather than two-dimensional. The second issue is that we used only one set of stimuli and one metaphor – buying paint from strangely shaped containers – and we could not generalize our findings based on that.

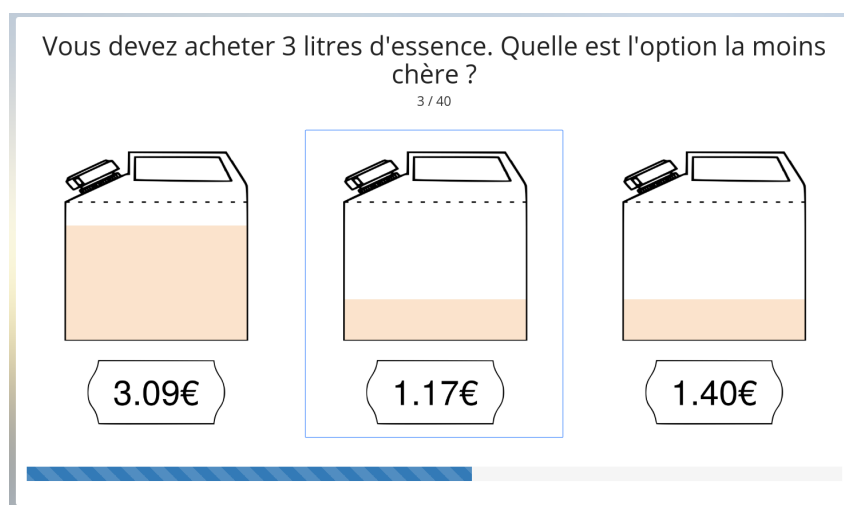


FIGURE 18: Stimuli used in the ongoing work on context effects

We used two sets of stimuli. The underlying task is the same – expenditure minimization in a spuriously complex scenario – but the metaphor used is one of buying gasoline. In the *visual* treatment, subjects see a jerry-can filled with gasoline (Figure 18) ; in the *numeric* treatment, the very same information is given in the form of numbers (here, liters). This second treatment is in principle more generalizable.

Provisional results show that the attraction effect is not robust to larger differences in prices and that it emerges for the most part only in the early stages of the search process. Consumers provisionally choose the target to avoid the decoy and then progressively switch from the target

to the competitor until their respective choice shares come to correspond to price differences only. Work is underway, a working paper is coming up by the end of 2017.

Overall, we have some (weak and preliminary) evidence that given enough time, the ADE dithers away. This, however, is valid in our experiment, in which preferences do not play a role, subjects are focused on the task and incentivized to put in their best effort. More extensive research is needed to ascertain the robustness of the ADE when more context or preferences are added to the picture.

Despite the relatively low importance given to the ADE and other context effects in economics, I do think that this line of research is promising. These effects have proved to be robust and evidence of the contrary, that we are starting to gather, might be of great importance in the consumer research and marketing literatures. Moreover, one of the reasons why ADE has been largely neglected by economists is that it goes against a fundamental axiom of decision theory, and no robust theory nor addition to existing models that can account for it has been proposed. Theories exist in psychology, but fail to account for all context effects, and in one case (Decision Field Theory, see Pettibone, 2012) predict that the ADE must *increase* in the time given to the subject, a prediction that is directly and strongly contradicted by our results. More work is needed to apply our measure to other context effects, and use the measurements – for the first time yielding an individual measure, a sign, and a magnitude – to build a better, stronger theory of context effects in multiple choice. If our new measure proves robust and gets some popularity, I think there is room for a revival of interest in context effects on the part of economists.

Risk taking

One possible reason for the low consistency across risk measures, and between these and self-reported measures might lie in the distinction between System One and System Two. Self-reported risk measures might be given by System One, while virtually all risk elicitation tasks mobilize at least partly System Two, since the decisions involve financial gains framed into lotteries, usually given to subjects in numerical formats.

The Fast&Slow elicitation mechanism could allow us to see if the low correlation exists from the start, from the first intuitive reply to a risk task, or else is something that happens in time, when subjects likely change their mind upon reflection.

Together with Antonio we ran a small, unincentivized pilot in Milan to explore this conjecture. We gave the subjects a battery of multiple choices extracted from the EG task – that is already framed as a single choice among multiple alternatives – and from a discretization and simplification of the BRET. Then, at the end, we asked the subjects to self-report their risk attitudes using both the SOEP risk question (Wagner, Frick, and Schupp, 2007) and the DOSPERT questionnaire (Blais and Weber, 2006).

Under the hypothesis that risk behavior changes when moving from System One to System Two, we should have been able to ascertain if subjects are intuitively risk averse, neutral or lover, and the direction of change once System Two gets into play. Moreover, we should have been able to measure correlation of both the fast and the slow measures with self-reported and real-world risk taking.

Unfortunately, provisional results are indeed extremely flat. Most subjects wait for a long time and choose only once, paying a price in terms of the possibility of being allocated a random choice but most probably not trusting themselves to submit a provisional fast choice given the 'difficult' stimuli. In the EG task subjects do seem to take more risk at the very beginning, but this is not significant; in the BRET, there is hardly any movement from nearly perfect risk neutrality all along. As a consequence, also correlations with the SOEP self-reported answer are flat in time, and hover around .3 for EG and .1 for the BRET.

Even if the first application of the F&S method to risk did not yield the expected results, I still think it of interest to investigate the reasons for the low consistency of estimates displayed by different risk elicitation tasks. Not because I think it possible to find one risk elicitation task that would beat the others, or because I find it very valuable to reconcile all the myriad ways in which tasks differ. But because given the abundance of data available and the keen interest of most economists and social scientists, risk is a perfect testing ground for new behavioral tools and theories.

Farther along the way

One of the key aspects of my career so far has been the urge to publish. Publish or perish, they say, and I wanted to be on the side of the living. Especially in the last two to three years, I have been dealing with several revisions, resubmissions, and rewritings of papers at the same time. I have been focused on polishing my past rather than being busy about my future. This work is not finished – several papers are still on my desk and drawers waiting to be brought to publication stage – but it is mostly done.

One of my dearest wishes from the year or two to come is to have the time to look a bit further. This means both reading papers in economics that I have been meaning to for a long time, *and* look in other disciplines for inspiration.

A first breath of fresh air came with the reflections that brought me to the INRA Young Researcher grant and to the projects described above. All of which are more or less ongoing. For some I have a clear path in mind, less for others, but they are all projects I started with a roadmap to results and publications.

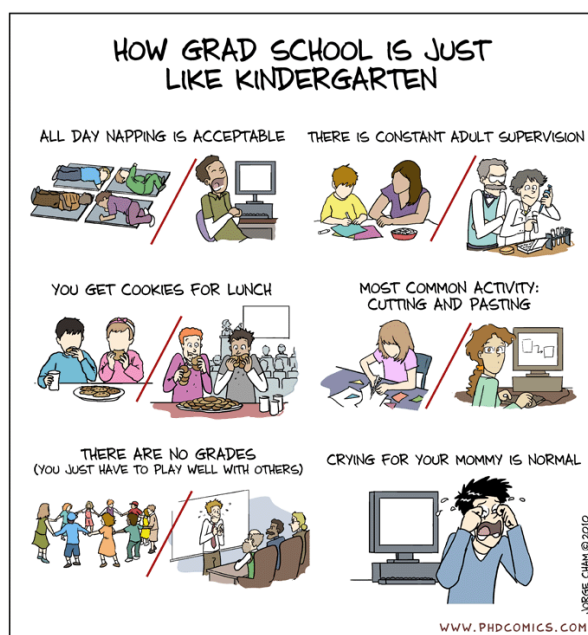
But I would like to allow myself the time to think and read a bit more. Finally read the last issues of the Journal of Economic Perspectives that lie around in my office. Read more cognitive psychology texts – and not just papers, but full length books, something I have been having no time for in the last four years at least. The price of this freedom is detail: I cannot give a precise roadmap here, because I don't know yet.

But I have some ideas. One comes from the recognition that framing and context effects are everywhere in experimental economics, yet they are treated amateurishly. There is no theory for when and why these effects come into play, how to avoid them or how to exploit them. Each experimenter knows how to avoid the most obvious pitfalls and can be more or less savvy about some of them. But I am unaware of rigorous efforts at understanding where they come from and what to do about it. I have been reading papers from cognitive psychology about *scripts* that dictate social acceptability standards, and I plan to read more of Simon's books on human problem solving and bounded rationality. Moreover, I have recently discovered the work of

Claude Shannon on the foundation of information technology, encoding and redundancy, and I found the idea that languages are largely made up of redundant information fascinating. This is just an intuition rather than anything else, but a possible rationalization of framing effects could lie in the fact that subjects try to minimize the information content of what they are doing – they look for actions that would give *order* to their experiences in the lab, and that are in a way *redundant*, do not bring along additional information.

While developing a whole new behavioral theory of context effects is beyond my reach, I hope to be able to transform some insights into tools to be used by experimental economists both in theoretical and applied work. This would combine my fascination for new approaches, methodological innovation, the possibility to run experiments and applications to various domains. It is a very large project; but it is worth a try.

What could I teach to students?



Having come full circle, it is time to ask the big question: am I ready to be a supervisor?

Having recently become a dad, I fear that the answer to this question is *not really*. For one is never fully prepared to be a dad until the son is born – and even then, it takes months and years and a lifetime of daddiness to become one. My mother told me, right before my son Pietro was born, that when expecting me she would worry about making the same mistakes her mother had made with her. A friend told her "*don't worry, you will not make them. You'll make others*". If I have to frankly answer the core question of my HDR, the most honest response I can come up with is that I'll try my best, I'll devote myself to it, and with the things I have learned, intelligent people around me, new mistakes and some luck I will little by little become a supervisor.

Truth be told, I have already started. Along the years I have been helping fellow PhD students, colleagues, master students, complete strangers over the Internet more than once. I am aware that these experiences might equip me for a real supervision as much as baby sitting a niece or a nephew equips you for parenthood – that is, not so much. But they did give me some background and taught me lessons.

In 2011 I was contacted by John Hey to help two PhD students – Giorgia Barboni, now post-doc at Princeton, and Tania Treibich, now assistant professor in Maastricht – for an experiment. The main deal was to write their experimental software since they were new to experiments, but the cooperation led to a bigger involvement. It is when sitting down to write the final software

that most details surface, and you have to deal with them. Their research was written up into Barboni and Treibich, 2013. My role was not major, but I count this as the first time I was directly involved in other student's work.

In 2012 I gave a talk in Göttingen about my Scrabble paper. I was later contacted by a PhD student there, Julia Brüggemann. She later became my co-author in a further Scrabble paper and I count her as my peer; but in a way our joint work felt a bit like a supervision, especially at the beginning. Again, the link was through the software and the Scrabble task I had developed. But again, the paper we later co-authored and that is part of Julia's PhD thesis was in a way the result of a sort of supervision. Julia's work benefited from mine, ideas were shaped in discussions, and she managed to get the ideas farther than I could have done.

I always try to be open for discussion and advice to anyone who cares to ask me. This means that I have been helping out students and colleagues from around the world, even if in several occasions I never met them in person (and most probably never will). I have been diffusing notes and practices about matching protocols in experiments to about half a dozen students now – it all started with Alexia Gaudeul pointing a student to me, and it continues. Since the BRET appeared, I routinely get contacted to give advice on how to adapt it to some special samples – most recently by Helene Willadsen, PhD student at Copenhagen, on how to adapt the BRET for children – or on how to port the software to other platforms – as Felix Holzmeister did, porting the BRET to OTree (Holzmeister and Pfurtscheller, 2016).

Most recently, I have been involved with the activities of the Graduate School (Ecole Doctorale) here in Grenoble. I acted as discussant for Lisa Charroin, a PhD student from Lyon, supervised by Christophe Bravard, a colleague at GAEL. I vouched for two PhD students at GAEL – Nicolas Pasquier et Bilel Rahali – to join the IMPRS Summer School in Jena (they survived it and loved it), I read and commented on their works. And I joined the *Comité de thèse*, the committee following closely a PhD student in France, of Nicolas Pasquier, who is supervised by my colleagues Alexis Garapin and Olivier Bonroy. After a first paper devoted to an IO theory of secret contracts, Nicolas is moving to experimental work and I am glad to be at his side.

Still, I ask myself what could I teach to PhD students.

I think the most important lesson is not to lose hope. It might not be the same for all PhD students, but the doctoral years are often difficult ones. A PhD is a lone endeavor, and for most people it is the first time they move from being very good at doing what they are told to study to being the ones that have to explore on their own. The fact that there is a very successful comic strip, PhDComics, whose main themes are frustration, solitude and cynicism says a lot about the psychological weight of a PhD.

Not losing hope works well also in another key domain: submissions. Having had my high share of rejections, like everyone, I learned how to fight back, when to accept criticism and go back to the drawing board, and the key insight: always resubmit. I also failed this – I have some papers in my drawers that I am unable or very slow in resubmitting. But the lesson learned is that you have to believe in your works, and believe hard, to overcome the odds.

Do not underestimate your work comes next. It helps to sit down and think about the reasons why a PhD started and the good things that will come out of it when it ends. I recently got to know that the University of Grenoble sponsors a course by a stand-up comedian to teach PhD students to present their thesis in 180 seconds. It is just a show, but out of it comes a great

lesson: students that joined in the program were urged to find the most important, sometimes hidden, beautiful and inspiring thing in their work, and communicate it to others. They had fun, and built stronger self confidence. The program is so successful that a PhD candidate from Grenoble won the French-wide competition, talking about how to better coat copper wires for electrons to sprint through. It does not seem the sexiest research topic, but she made her sound it like that. It's an amazing achievement and a message that is important to give to PhD and researchers in general: *this is cool; you matter*. You can see her her 180-second speech here (in French): https://www.youtube.com/watch?v=-i_3ALztV4s.

Last lesson: talk a lot about your work, and do not be afraid to ask questions. Present a lot. Through presentations a researcher can try to put himself in the shoes of his readers and listeners, and hence get a new perspective on his work. Ask questions. Questions turn into answers, but also involve other people in your work, and you always learn from them.

Then, of course, I think I could teach them something about experimental economics. Several existing projects seem to have the potential for a student to join me in my efforts, and new laboratories and equipments are about to be added to the facilities existing in Grenoble that could make further research possible. The Fast&Slow projects are all on their way, but they might open up avenues of research. In the food domain, lots is still to be done to enlarge the set of methods applied to food choices. A new laboratory with 40 seats, eye scanners and other physiological measuring equipment will be ready early in 2019, making it possible to run experiments merging physical, cognitive and economic measures; moreover, recently on the Grenoble campus an interdisciplinary group on food research was formed, opening fascinating possibilities for cross-fertilization. In the domain of context effects, the main aim of my future research will be to experiment on other context effect other than the ADE, and work towards a unified explanation, from which possibly to build a theory. The biggest goal of new risk research would be to look for a unified measure able to robustly predict risk attitudes – or, if that fails, at least to understand how and why our existing measures fail.

But most than everything, I see supervision as a chance to learn new things together with bright and enthusiastic people. I'm looking forward to it.

Bibliography

- Akerlof, George A and Rachel E Kranton (2000). "Economics and identity". In: *Quarterly Journal of Economics* 115.3, pp. 715–753.
- Allais, Maurice (1953). "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine". In: *Econometrica* 31, 503–546.
- Anderson, Lisa and Charles Holt (1997). "Information Cascades in the Laboratory". In: *American Economic Review* 87.5, pp. 847–62.
- Anderson, Lisa R. and Jennifer M. Mellor (2009). "Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure". In: *Journal of Risk and Uncertainty* 39.2, pp. 137–160.
- Anderson, Philip W et al. (1972). "More is different". In: *Science* 177.4047, pp. 393–396.
- Arce Salazar, Helen and Leon Oerlemans (2016). "Do We Follow the Leader or the Masses? Antecedents of the Willingness to Pay Extra for Eco-Products". In: *Journal of Consumer Affairs* 50.2, pp. 286–314.
- Arrow, Kenneth J. (1962). "The rate and Direction of Incentive Activity. Economics and Social Factors". In: ed. by R.R. Nelson. Princeton University Press. Chap. Economic Welfare and the Allocation of Resources for Invention.
- Asch, S. E. (1955). "Opinions and Social Pressure". In: *Scientific American* 193, p. 3535.
- Asch, S.E. (1951). "Groups, leadership and men". In: ed. by H. Guetzkow. Pittsburgh, PA: Carnegie Press. Chap. Effects of group pressure on the modification and distortion of judgments, pp. 177–190.
- Ashton, Lydia (2015). "Hunger Games: Does Hunger Affect Time Preferences?" In: Available at SSRN 2538740.
- Axelrod, Robert (1984). *The Evolution of Cooperation*. Basic Books, New York.
- Banerjee, Abhijit V. (1992). "A Simple Model of Herd Behavior". In: *Quarterly Journal of Economics* 107.3, pp. 797–817.
- Barber, Brad M and Terrance Odean (2001). "Boys will be boys: Gender, overconfidence, and common stock investment". In: *The quarterly journal of economics* 116.1, pp. 261–292.
- Barboni, Giorgia and Tania Treibich (2013). *Risk aversion and signalling in single and multiple-bank lending*. GREDEG Working Papers 2013-28. Groupe de REcherche en Droit, Economie, Gestion (GREDEG CNRS), University of Nice Sophia Antipolis. URL: <https://ideas.repec.org/p/gre/wpaper/2013-28.html>.
- Bechtold, Stefan, Christopher Buccafusco, and Christopher Jon Springman (2015). "Innovation heuristics: experiments on sequential creativity in intellectual property". In: *Ind. LJ* 91, p. 1251.
- Belleflamme, Paul, Thomas Lambert, and Armin Schwienbacher (2014). "Crowdfunding: Tapping the right crowd". In: *Journal of Business Venturing* 29, pp. 585–609.
- Benassi, Victor A., Paul D. Sweeney, and Gregg E. Drevno (1979). "Mind over matter: Perceived success at psychokinesis". In: *Journal of Personality and Social Psychology* 37.8, pp. 1377–1386.

- Benkler, Yochai (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press. ISBN: 0300110561.
- Bessen, James and Eric Maskin (2009). "Sequential innovation, patents, and imitation". In: *The RAND Journal of Economics* 40.4, pp. 611–635. ISSN: 07416261. DOI: [10.1111/j.1756-2171.2009.00081.x](https://doi.org/10.1111/j.1756-2171.2009.00081.x).
- Bialkova, Svetlana and Hans van Trijp (2010). "What determines consumer attention to nutrition labels?" In: *Food Quality and Preference* 21.8. Eighth Pangborn Sensory Science Symposium, pp. 1042–1051. ISSN: 0950-3293. DOI: <http://dx.doi.org/10.1016/j.foodqual.2010.07.001>. URL: <http://www.sciencedirect.com/science/article/pii/S0950329310001199>.
- Bicchieri, Cristina and Erte Xiao (2009). "Do the right thing: but only if others do so". In: *Journal of Behavioral Decision Making* 22.2, pp. 191–208.
- Blais, Anne-Renee and Elke U. Weber (2006). "A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations". In: *Judgment and Decision Making* 1, pp. 33–47.
- Boldrin, Michele and David K. Levine (2008a). *Against intellectual monopoly*. Ed. by. New York, NY [u.a.]: Cambridge Univ. Press. VIII, 298. ISBN: 978-0-521-87928-6.
- (2008b). "Perfectly competitive innovation". In: *Journal of Monetary Economics* 55.3, pp. 435–453. URL: <http://ideas.repec.org/a/eee/moneco/v55y2008i3p435-453.html>.
- Bolton, Gary E. and Rami Zwick (1995). "Anonymity versus Punishment in Ultimatum Bargaining". In: *Games and Economic Behavior* 10.1, pp. 95–121. ISSN: 0899-8256.
- Bonroy, Olivier and Christos Constantatos (2015). "On the Economics of Labels: How Their Introduction Affects the Functioning of Markets and the Welfare of All Participants". In: *American Journal of Agricultural Economics* 97.1, pp. 239–259. URL: <https://ideas.repec.org/a/oup/ajagec/v97y2015i1p239-259..html>.
- Borgmeier, Ingrid and Joachim Westenhoefer (2009). "Impact of different food label formats on healthiness evaluation and food choice of consumers: a randomized-controlled study". In: *BMC Public Health* 9.
- Botzen, W.J.W. and J.C.J.M. van den Bergh (2012). "Risk attitudes to low-probability climate change risks: WTP for flood insurance". In: *Journal of Economic Behavior & Organization* 82.1, pp. 151–166. ISSN: 0167-2681. DOI: [10.1016/j.jebo.2012.01.005](https://doi.org/10.1016/j.jebo.2012.01.005).
- Bravo, Giangiacomo and Flaminio Squazzoni (2013). "Exit, punishment and rewards in commons dilemmas: an experimental study". In: *PloS one* 8.8, e69871. DOI: [10.1371/journal.pone.0069871](https://doi.org/10.1371/journal.pone.0069871). URL: <http://dx.plos.org/10.1371/journal.pone.0069871.g003> (visited on 11/04/2014).
- Bruner, David (2009). "Changing the probability versus changing the reward". In: *Experimental Economics* 12.4, pp. 367–385.
- Brüggemann, Julia et al. (2016). "Intellectual property rights hinder sequential innovation. Experimental evidence". In: *Research Policy* 45.10, pp. 2054–2068. ISSN: 0048-7333. DOI: <http://dx.doi.org/10.1016/j.respol.2016.07.008>. URL: <http://www.sciencedirect.com/science/article/pii/S0048733316301159>.
- Buccafusco, Christopher J. and Christopher Jon Sprigman (2010). "Valuing Intellectual Property: An Experiment". In: *Cornell Law Review* 91. URL: <http://dx.doi.org/10.2139/ssrn.1568962>.
- (2011). "The Creativity Effect". In: *University of Chicago Law Review* 78, p. 31. URL: <http://ssrn.com/abstract=1647009>.

- Buchanan, Joy A. and Bart J. Wilson (2012). *An Experiment on Protecting Intellectual Property*. Working Papers 12-09. Chapman University, Economic Science Institute. URL: <http://ideas.repec.org/p/chu/wpaper/12-09.html>.
- Byrnes, James P, David C Miller, and William D Schafer (1999). "Gender differences in risk taking: A meta-analysis." In: *Psychological bulletin* 125.3, p. 367.
- Camerer, Colin, George Loewenstein, and Drazen Prelec (2005). "Neuroeconomics: How neuroscience can inform economics". In: *Journal of Economic Literature* 43.1, pp. 9–64.
- Cantner, Uwe et al. (2009). "Competition In Product Design: An Experiment Exploring Innovation Behavior". In: *Metroeconomica* 60.4, pp. 724–752. URL: <http://ideas.repec.org/a/bla/metroe/v60y2009i4p724-752.html>.
- Caplin, Andrew, Mark Dean, and Daniel Martin (2011). "Search and Satisficing". In: *American Economic Review* 101.7, pp. 2899–2922.
- Charness, Gary and Uri Gneezy (2010). "Portfolio Choice And Risk Attitudes: An Experiment". In: *Economic Inquiry* 48.1, pp. 133–146.
- (2012). "Strong Evidence for Gender Differences in Risk Taking". In: *Journal of Economic Behavior & Organization* 83.1, pp. 50–58. DOI: [10.1016/j.jebo.2011.06.007](https://doi.org/10.1016/j.jebo.2011.06.007).
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates. ISBN: 9780805802832.
- Colombo, Massimo G., Chiara Franzoni, and Cristina Rossi-Lamastra (2015). "Internal Social Capital and the Attraction of Early Contributions in Crowdfunding". In: *Entrepreneurship Theory and Practice* 39.1, pp. 75–100.
- Crosetto, Paolo (2010). "To patent or not to patent : a pilot experiment on incentives to copyright in a sequential innovation setting". In: *Department of Economics, University of Milan*.
- Crosetto, Paolo and Antonio Filippin (2013). "The 'Bomb' Risk Elicitation Task". In: *Journal of Risk and Uncertainty* 47.1, pp. 31–65.
- (2015). "A theoretical and experimental appraisal of four risk elicitation methods". English. In: *Experimental Economics*, pp. 1–29. ISSN: 1386-4157. DOI: [10.1007/s10683-015-9457-9](https://doi.org/10.1007/s10683-015-9457-9). URL: <http://dx.doi.org/10.1007/s10683-015-9457-9>.
- (2017a). *Safe options induce gender differences in risk attitudes*. Working Papers 05/2017. GAEL Working Paper. URL: <https://ideas.repec.org/p/hal/wpaper/halshs-01529643.html>.
- (2017b). "The Sound of Others: Surprising Evidence of Conformist Behavior". In: *Southern Economic Journal* 83.4, pp. 1038–1051. ISSN: 2325-8012. DOI: [10.1002/soej.12186](https://doi.org/10.1002/soej.12186). URL: <http://dx.doi.org/10.1002/soej.12186>.
- Crosetto, Paolo, Antonio Filippin, and Janna Heider (2015). "A Study of Outcome Reporting Bias Using Gender Differences in Risk Attitudes". In: *CESifo Economic Studies* 61.1, pp. 239–262.
- Crosetto, Paolo and Alexia Gaudeul (2016). "A monetary measure of the strength and robustness of the attraction effect". In: *Economics Letters* 149, pp. 38–43. ISSN: 0165-1765. DOI: [10.1016/j.econlet.2016.09.031](https://doi.org/10.1016/j.econlet.2016.09.031). URL: <http://www.sciencedirect.com/science/article/pii/S0165176516303901>.
- (2017). "Choosing not to compete: Can firms maintain high prices by confusing consumers?" In: *Journal of Economics & Management Strategy*. DOI: [10.1111/jems.12212](https://doi.org/10.1111/jems.12212). URL: <http://onlinelibrary.wiley.com/doi/10.1111/jems.12212/abstract>.
- Crosetto, Paolo, Laurent Muller, and Bernard Ruffieux (2016a). "Helping consumers with a front-of-pack label: Numbers or colors?: Experimental comparison between Guideline Daily Amount and Traffic Light in a diet-building exercise". In: *Journal of Economic Psychology*

55. Special issue on Food consumption behavior: Economic and psychological perspectives, pp. 30–50. ISSN: 0167-4870. DOI: <http://dx.doi.org/10.1016/j.joep.2016.03.006>.
- Crosetto, Paolo, Laurent Muller, and Bernard Ruffieux (2016b). “Réponses des consommateurs à trois systèmes d’étiquetage nutritionnel face avant”. In: *Cahiers de Nutrition et de Diététique* 5011.3, pp. 111–160. ISSN: 0007-9960. DOI: <http://dx.doi.org/10.1016/j.cnd.2016.04.002>. URL: [http://www.sciencedirect.com/science/article/pii/S0007-9960\(16\)30034-7](http://www.sciencedirect.com/science/article/pii/S0007-9960(16)30034-7).
- Crosetto, Paolo and Tobias Regner (2014). *Crowdfunding: Determinants of success and funding dynamics*. Jena Economic Research Papers 2014-035. Friedrich-Schiller-University Jena. URL: <https://ideas.repec.org/p/jrp/jrpwrp/2014-035.html>.
- Crosetto, Paolo et al. (2012). *Motives of Sanctioning: Equity and Emotions in a Public Good Experiment with Punishment*. Jena Economic Research Papers 2012-046. Friedrich-Schiller-University Jena. URL: <https://ideas.repec.org/p/jrp/jrpwrp/2012-046.html>.
- Crosetto, Paolo et al. (2017). “Modification des achats alimentaires en réponse à cinq logos nutritionnels”. In: *Cahiers de Nutrition et de Diététique* 5262.3, pp. 107–159. ISSN: 0007-9960. DOI: <http://dx.doi.org/10.1016/j.cnd.2017.04.002>. URL: [http://www.sciencedirect.com/science/article/pii/S0007-9960\(17\)30070-6](http://www.sciencedirect.com/science/article/pii/S0007-9960(17)30070-6).
- Crosen, Rachel and Uri Gneezy (2009). “Gender Differences in Preferences”. In: *Journal of Economic Literature* 47.2, pp. 448–74.
- Darmon, Nicole et al. (2014). “Food price policies improve diet quality while increasing socio-economic inequalities in nutrition”. In: *Int J Behav Nutr Phys Act* 11, p. 66.
- Dave, Chetan et al. (2010). “Eliciting risk preferences: When is simple better?” In: *Journal of Risk and Uncertainty* 41.3, pp. 219–243.
- Deck, Cary et al. (2013). “A failed attempt to explain within subject variation in risk taking behavior using domain specific risk attitudes”. In: *Journal of Economic Behavior & Organization* 87, pp. 1–24.
- Dimmig, Anne-Kathrin and Mathias Erlei (2012). *Quasi-rational R&D behavior in an environment with fundamental uncertainty*. Tech. rep. TUC Working Papers in Economics.
- Dong, Xiao Yuan and Gregory K. Dow (1993). “Does Free Exit Reduce Shirking in Production Teams?” In: *Journal of Comparative Economics* 17.2, pp. 472–484. ISSN: 0147-5967. DOI: [10.1006/jcec.1993.1037](https://doi.org/10.1006/jcec.1993.1037).
- Doyle, Joseph, Erich Muehlegger, and Krislert Samphantharak (2010). “Edgeworth cycles revisited”. In: *Energy Economics* 32.3, pp. 651–660. ISSN: 0140-9883. DOI: [10.1016/j.eneco.2009.09.001](https://doi.org/10.1016/j.eneco.2009.09.001). URL: <http://www.sciencedirect.com/science/article/pii/S0140988309001571> (visited on 12/14/2015).
- Drichoutis, Andreas C, Panagiotis Lazaridis, and Rodolfo M Nayga (2009). “Would consumers value food-away-from-home products with nutritional labels?” In: *Agribusiness* 25.4, pp. 550–575.
- Drichoutis, Andreas C., Rodolfo M. Nayga, and Panagiotis Lazaridis (2011). “Nutritional Labeling”. In: ed. by Jayson L. Lusk, Jutta Roosen, and Jason F. Shogren. Oxford University Press, Oxford. Chap. 20, pp. 520–545.
- Drichoutis, Andreas C, Rodolfo M Nayga, and Panagiotis Lazaridis (2012). “Food away from home expenditures and obesity among older Europeans: are there gender differences?” In: *Empirical Economics* 42.3, pp. 1051–1078.

- Ducrot, Pauline et al. (2015). "Objective Understanding of Front-of-Package Nutrition Labels among Nutritionally At-Risk Individuals". In: *Nutrients* 7.8, pp. 7106–7125. ISSN: 2072-6643. DOI: [10.3390/nu7085325](https://doi.org/10.3390/nu7085325). URL: <http://www.mdpi.com/2072-6643/7/8/5325>.
- Eagly, Alice H. and Linda Carli (1981). "Sex of researchers and sex-typed communications as determinants of sex differences in influenceability: A meta-analysis of social influence studies". In: *Psychological Bulletin* 90, pp. 1–20.
- Eckel, Catherine C. and Philip J. Grossman (2008a). "Chapter 113 Men, Women and Risk Aversion: Experimental Evidence". In: *Handbook of Experimental Economics Results 1*. Ed. by Charles R. Plott and Vernon L. Smith, pp. 1061–1073. ISSN: 1574-0722. DOI: [10.1016/S1574-0722\(07\)00113-8](https://doi.org/10.1016/S1574-0722(07)00113-8).
- (2008b). "Forecasting risk attitudes: An experimental study using actual and forecast gamble choices". In: *Journal of Economic Behavior & Organization* 68.1, pp. 1–17. ISSN: 0167-2681. DOI: [10.1016/j.jebo.2008.04.006](https://doi.org/10.1016/j.jebo.2008.04.006).
- Ederer, Florian and Gustavo Manso (2013). "Is pay for performance detrimental to innovation?" In: *Management Science* 59.7, pp. 1496–1513.
- Ellison, Glenn (2006). "Bounded rationality in industrial organization". In: *Advances in economics and econometrics: theory and applications. Ninth World Congress of the Econometric Society*. Ed. by Whitney K. Newey Torsten Persson Richard Blundell. Vol. 2. Cambridge University Press.
- Filippin, Antonio and Paolo Crosetto (2016a). *A reconsideration of gender differences in risk attitudes*. Tech. rep. 11, pp. 3138–3160.
- (2016b). "Click'n'Roll: No Evidence of Illusion of Control". In: *De Economist* 164.3, pp. 281–295. ISSN: 1572-9982. DOI: [10.1007/s10645-016-9282-3](https://doi.org/10.1007/s10645-016-9282-3). URL: <https://doi.org/10.1007/s10645-016-9282-3>.
- Frederick, Shane, Leonard Lee, and Ernest Baskin (2014). "The Limits of Attraction". In: *Journal of Marketing Research* 51.4, pp. 487–507.
- Gaudeul, Alexia, Paolo Crosetto, and Gerhard Riener (2017). "Better stuck together or free to go? Of the stability of cooperation when individuals have outside options". In: *Journal of Economic Psychology* 59, pp. 99–112. ISSN: 0167-4870. DOI: [http://dx.doi.org/10.1016/j.joep.2017.01.005](https://doi.org/10.1016/j.joep.2017.01.005). URL: <http://www.sciencedirect.com/science/article/pii/S016748701630099X>.
- Gerber, Elizabeth M and Julie Hui (2013). "Crowdfunding: Motivations and deterrents for participation". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 20.6, p. 34.
- Gigerenzer, Gerd and Daniel G. Goldstein (1999). "Betting on one good reason: The take the best heuristic". In: *Simple heuristics that make us smart*. Ed. by Gerd Gigerenzer, Peter M. Todd, and The ABC Research Group. Oxford University Press, pp. 75–95.
- Glimcher, Paul W and Ernst Fehr (2013). *Neuroeconomics: Decision making and the brain*. Academic Press.
- Gneezy, Uri, Kenneth L. Leonard, and John A. List (2009). "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society". In: *Econometrica* 77.5, pp. 1637–64.
- Gneezy, Uri, Muriel Niederle, and Aldo Rustichini (2003). "Performance in Competitive Environments: Gender Differences". In: *Quarterly Journal of Economics* CXVIII, pp. 1049–1074.
- Gneezy, Uri and Jan Potters (1997). "An Experiment on Risk Taking and Evaluation Periods". In: *The Quarterly Journal of Economics* 112.2, pp. 631–45.
- Goeree, Jacob K. and Leet Yariv (2015). "Conformity in the lab". In: *Journal of the Economic Science Association*, pp. 1–22.

- Gorton, Delvina (2007). "Nutrition labelling-Update of scientific evidence on consumer use and understanding of nutrition labels and claims". In: *Retrieved May 14*, p. 2011.
- Grunert, Klaus G. and Josephine M. Wills (2007). "A review of European research on consumer response to nutrition information on food labels". English. In: *Journal of Public Health* 15.5, pp. 385–399. ISSN: 0943-1853. DOI: [10.1007/s10389-007-0101-9](https://doi.org/10.1007/s10389-007-0101-9).
- Güth, Werner (1995). "On ultimatum bargaining experiments — A personal review". In: *Journal of Economic Behavior & Organization* 27.3, pp. 329–344.
- Güth, Werner and Martin G. Kocher (2014). "More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature". In: *Journal of Economic Behavior & Organization* 108, pp. 396–409.
- Haney, Craig, Curtis Banks, and Philip Zimbardo (1973). "A study of prisoners and guards in a simulated prison". In: *Naval Research Reviews* 30, pp. 4–17.
- Harbaugh, W.T., K. Krause, and L. Vesterlund (2010). "The Fourfold Pattern of Risk Attitudes in Choice and Pricing Tasks". In: *The Economic Journal* 120.545, pp. 595–611.
- Harrison, Glenn W. (2008). "Neuroeconomics: A Critical Reconsideration". In: *Economics and Philosophy* 24.3, pp. 303–344.
- Hawley, Kristy L et al. (2013). "The science on front-of-package food labels". In: *Public Health Nutrition* 16, pp. 430–9.
- Heller, Michael A. and Rebecca S. Eisenberg (1998). "Can Patents Deter Innovation? The Anti-commons in Biomedical Research". In: *Science* 280.5364, pp. 698–701. DOI: [10.1126/science.280.5364.698](https://doi.org/10.1126/science.280.5364.698). eprint: <http://www.sciencemag.org/cgi/reprint/280/5364/698.pdf>. URL: <http://www.sciencemag.org/cgi/content/abstract/280/5364/698>.
- Herne, Kaisa (1999). "The Effects of Decoy Gambles on Individual Choice". English. In: *Experimental Economics* 2.1, pp. 31–40. ISSN: 1386-4157. DOI: [10.1023/A:1009925731240](https://doi.org/10.1023/A:1009925731240). URL: <http://dx.doi.org/10.1023/A%3A1009925731240>.
- Hey, John D and Chris Orme (1994). "Investigating Generalizations of Expected Utility Theory Using Experimental Data". In: *Econometrica* 62.6, pp. 1291–1326.
- Hirschman, Albert O. (1970). *Exit, Voice and Loyalty: Responses to Decline in Firms, Organizations, and States*. Harvard University Press.
- Holt, C.A. and S.K. Laury (2002). "Risk aversion and incentive effects". In: *American Economic Review* 92.5, pp. 1644–1655.
- Holt, Charles A. (1993). "Industrial organization: A survey of laboratory research". In: *The Handbook of Experimental Economics* 349, pp. 402–03. URL: <http://www.people.virginia.edu/~cah2k/iosurvtr.pdf> (visited on 11/03/2014).
- Holzmeister, Felix and Armin Pfurtscheller (2016). "oTree: The "bomb" risk elicitation task". In: *Journal of Behavioral and Experimental Finance* 10, pp. 105–108. ISSN: 2214-6350. DOI: [http://dx.doi.org/10.1016/j.jbef.2016.03.004](https://doi.org/10.1016/j.jbef.2016.03.004). URL: <http://www.sciencedirect.com/science/article/pii/S2214635016300120>.
- Huber, Joel, John W. Payne, and Christopher Puto (1982). "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis". In: *Journal of Consumer Research* 9.1, pp. 90–98.
- Huber, Joel, John W. Payne, and Christopher P. Puto (2014). "Let's Be Honest About the Attraction Effect". In: *Journal of Marketing Research* 51.4, pp. 520–525. DOI: [10.1509/jmr.14.0208](https://doi.org/10.1509/jmr.14.0208).
- Huber, Joel and Christopher Puto (1983). "Market Boundaries and Product Choice: Illustrating Attraction and Substitution Effects". In: *Journal of Consumer Research* 10.1, pp. 31–41.

- Isaac, R.M. and D. James (2000). "Just who are you calling risk averse?" In: *Journal of Risk and Uncertainty* 20.2, pp. 177–187.
- Jimenez, Francisca Jimenez (2003). *Does Familiar Environment Affect Individual Risk Attitudes? Olive-oil Producer vs. no-producer Households*. Economic Working Papers at Centro de Estudios Andaluces E2003/28. Centro de Estudios Andaluces.
- Kahneman, Daniel (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux. ISBN: 9780374275631 0374275637.
- Kalaycı, Kenan (2011). *Price complexity and buyer confusion in markets*. Netspar Discussion Paper 10/2011-082.
- Kalaycı, Kenan and Jan Potters (2011). "Buyer confusion and market prices". en. In: *International Journal of Industrial Organization* 29.1, pp. 14–22. ISSN: 01677187. DOI: 10.1016/j.ijindorg.2010.06.004. URL: <http://linkinghub.elsevier.com/retrieve/pii/S0167718710000949> (visited on 11/03/2014).
- Kalaycı, Kenan and Marta Serra-Garcia (2015). "Complexity and biases". en. In: *Experimental Economics*, pp. 1–20. ISSN: 1386-4157, 1573-6938. DOI: 10.1007/s10683-015-9434-3. URL: <http://link.springer.com/article/10.1007/s10683-015-9434-3> (visited on 03/11/2015).
- Kelly, Bridget et al. (2009). "Consumer testing of the acceptability and effectiveness of front-of-pack food labelling systems for the Australian grocery market". In: *Health Promotion International* 24.2, pp. 120–129. DOI: 10.1093/heapro/dap012. eprint: <http://heapro.oxfordjournals.org/content/24/2/120.full.pdf+html>. URL: <http://heapro.oxfordjournals.org/content/24/2/120.abstract>.
- Klein, Stanley A. (2001). "Measuring, estimating, and understanding the psychometric function: A commentary". In: *Perception & Psychophysics* 63.8, pp. 1421–1455.
- Kocher, Martin G., Julius Pahlke, and Stefan T. Trautmann (2013). "Tempus Fugit: Time Pressure in Risky Decisions". In: *Management Science* 59.10, pp. 2380–2391. DOI: 10.1287/mnsc.2013.1711. eprint: <http://pubsonline.informs.org/doi/pdf/10.1287/mnsc.2013.1711>.
- Kuppuswamy, Venkat and Barry L Bayus (2015). "Crowdfunding creative ideas: The dynamics of project backers in Kickstarter". In:
- Langer, Ellen J (1975). "The illusion of control". In: *Journal of Personality and Social Psychology* 32.2, pp. 311–328.
- Lejuez, CW et al. (2002). "Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART)." In: *Journal of Experimental Psychology: Applied* 8.2, p. 75.
- Lessig, Lawrence (2001). *The Future of Ideas - The fate of the Commons in a connected world*. Random House, New York.
- Li, King King (2011). "Preference towards control in risk taking: Control, no control, or randomize?" In: *Journal of Risk and Uncertainty* 43 (1), pp. 39–63. ISSN: 0895-5646. DOI: 10.1007/s11166-011-9122-4.
- Lin, Justin Yifu (1990). "Collectivization and China's Agricultural Crisis in 1959-1961". English. In: *Journal of Political Economy* 98.6, pp. 1228–1252. ISSN: 00223808.
- Lunn, Pete and Jason J. Somerville (2015). *Surplus Identification with Non-Linear Returns*. Papers WP522. Economic and Social Research Institute (ESRI).
- Lusk, Jayson L (2003). "Effects of cheap talk on consumer willingness-to-pay for golden rice". In: *American Journal of Agricultural Economics* 85.4, pp. 840–856.
- Lusk, Jayson L. and Keith H. Coble (2005). "Risk Perceptions, Risk Preference, and Acceptance of Risky Food". In: *American Journal of Agricultural Economics* 87.2, pp. 393–405. DOI: 10.

- 1111/j.1467-8276.2005.00730.x. eprint: <http://ajae.oxfordjournals.org/content/87/2/393.full.pdf+html>.
- Lusk, Jayson L, F Bailey Norwood, and J Ross Pruitt (2006). "Consumer demand for a ban on antibiotic drug use in pork production". In: *American Journal of Agricultural Economics* 88.4, pp. 1015–1033.
- Lusk, Jayson L et al. (2006). "Consumer behavior, public policy, and country-of-origin labeling". In: *Review of agricultural economics* 28.2, pp. 284–292.
- MacLeod, W. Bentley (1988). "Equity, efficiency and incentives in cooperative teams". In: *Advances in the economic analysis of participatory and labor-managed firms*. Ed. by D. C. Jones and J. Svejnar. Vol. 3. JAI Press, pp. 5–23.
- (1993). "The Role of Exit Costs in the Theory of Cooperative Teams: A Theoretical Perspective". In: *Journal of Comparative Economics* 17.2, pp. 521–529. ISSN: 0147-5967. DOI: [10.1006/jcec.1993.1040](https://doi.org/10.1006/jcec.1993.1040).
- Meloso, Debrah, Jernej Copic, and Peter Bossaerts (2009). "Promoting intellectual discovery: patents versus markets". In: *Science* 323.5919, pp. 1335–1339.
- Menkhoff, Lukas and Sahra Sakha (2017). "Estimating risky behavior with multiple-item risk measures". In: *Journal of Economic Psychology* 59, pp. 59–86. ISSN: 0167-4870. DOI: <http://dx.doi.org/10.1016/j.joep.2017.02.005>. URL: <http://www.sciencedirect.com/science/article/pii/S0167487016304196>.
- Moeser, Anke et al. (2010). "Simplified nutrient labelling: consumers' perceptions in Germany and Belgium". English. In: *Journal fuer Verbraucherschutz und Lebensmittelsicherheit* 5.2, pp. 169–180. ISSN: 1661-5751. DOI: [10.1007/s00003-009-0531-0](https://doi.org/10.1007/s00003-009-0531-0).
- Møllgaard, H. Peter and Per Baltzer Overgaard (2001). "Market Transparency and Competition Policy". In: *Rivista di Politica Economica* 91.4, pp. 11–64. URL: <http://ideas.repec.org/a/rpo/ripec/v91y2001i4p11-64.html>.
- Mulder, Laetitia B. et al. (2006). "Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas". en. In: *Journal of Experimental Social Psychology* 42.2, pp. 147–162. ISSN: 00221031. DOI: [10.1016/j.jesp.2005.03.002](https://doi.org/10.1016/j.jesp.2005.03.002). URL: <http://dx.doi.org/10.1016/j.jesp.2005.03.002>.
- Muller, Laurent and Bernard Ruffieux (2012). "Modification des achats en réponse à l'apposition de différents logos d'évaluation nutritionnelle sur la face avant des emballages". In: *Cahiers de Nutrition et de Diététique* 47.4, pp. 171–182.
- Muller, Laurent et al. (2017). "Distributional impacts of fat taxes and thin subsidies". In: *The Economic Journal*.
- Nelson, Julie A. (2013). "Not-So-Strong Evidence for Gender Differences in Risk Taking". In: *Feminist Economics* 0.0, pp. 1–29. DOI: [10.1080/13545701.2015.1057609](https://doi.org/10.1080/13545701.2015.1057609).
- (2014). "Are Women Really More Risk-Averse than Men? A Re-Analysis of the Literature Using Expanded Methods". In: *Journal of Economic Surveys* 29.3, pp. 566–585. ISSN: 1467-6419. DOI: [10.1111/joes.12069](https://doi.org/10.1111/joes.12069).
- Nielsen, Thea, Alwin Keil, and Manfred Zeller (2013). "Assessing farmers' risk preferences and their determinants in a marginal upland area of Vietnam: a comparison of multiple elicitation techniques". In: *Agricultural Economics* 44.3, pp. 255–273.
- Orbell, John M., Peregrine Schwartz-Shea, and Randy T. Simmons (1984). "Do Cooperators Exit More Readily than Defectors?" In: *The American Political Science Review* 78.1, pp. 147–162. ISSN: 00030554. DOI: [10.2307/1961254](https://doi.org/10.2307/1961254). URL: <http://www.jstor.org/stable/1961254?origin=crossref> (visited on 11/04/2014).

- Perloff, Jeffrey M. and Steven C. Salop (1985). "Equilibrium with Product Differentiation". In: *The Review of Economic Studies* 52.1, pp. 107–120. ISSN: 00346527. DOI: [10.2307/2297473](https://doi.org/10.2307/2297473). URL: <http://restud.oxfordjournals.org/lookup/doi/10.2307/2297473> (visited on 11/03/2014).
- Pettibone, Jonathan C (2012). "Testing the effect of time pressure on asymmetric dominance and compromise decoys in choice". In: *Judgment and Decision Making* 7.4, p. 513.
- Poon, Derrick (2011). *Re-examining Behavioural Anomalies: Preference Reversal and Illusion of Control*. Unpublished Dissertation. School of Economics, The University of New South Wales.
- Presson, Paul K. and Victor A. Benassi (1996). "Illusion of control: A meta-analytic review". In: *Journal of Social Behavior and Personality* 11.3, pp. 493–510.
- Quiggin, John (1982). "A theory of anticipated utility". In: *Journal of Economic Behavior & Organization* 3.4, pp. 323–343.
- Rayner, M, Peter Scarborough, and Tim Lobstein (2009). "The UK Ofcom Nutrient Profiling Model: Defining 'healthy' and 'unhealthy' foods and drinks for TV advertising to children". In: *London: OfCom*.
- Regner, Tobias and Paolo Crosetto (2017). *The experience matters: participation-related rewards increase the success chances of crowdfunding campaigns*. Tech. rep. 04/2017. Gael Working Paper O4/2017. GAEL Working Papers. URL: <https://hal.archives-ouvertes.fr/hal-01527150>.
- Reutskaja, Elena et al. (2011). "Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study". In: *American Economic Review* 101.2, pp. 900–926.
- Reynaud, Arnaud and Stéphane Couture (2012). "Stability of risk preference measures: results from a field experiment on French farmers". In: *Theory and Decision* 73.2, pp. 203–221.
- Rijt, Arnout van de et al. (2014). "Field experiments of success-breeds-success dynamics". In: *Proceedings of the National Academy of Sciences* 111.19, pp. 6934–6939.
- Roosen, Jutta, Jayson L Lusk, and John A Fox (2003). "Consumer demand for and attitudes toward alternative beef labeling strategies in France, Germany, and the UK". In: *Agribusiness* 19.1, pp. 77–90.
- Roth, Alvin E. et al. (1991). "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study". In: *American Economic Review* 81.5, pp. 1068–95. URL: <http://EconPapers.repec.org/RePEc:aea:aecrev:v:81:y:1991:i:5:p:1068-95>.
- Ruffieux, Bernard et al. (2014). *Projet Lab2Green. Une étude expérimentale de l'impact sur les choix des consommateurs des différents systèmes d'évaluation environnemental apposés en face avant des aliments. Rapport scientifique*. Contract. Ministère de l'Ecologie, du Développement Durable et de l'Energie, 62 p. URL: <https://hal.archives-ouvertes.fr/hal-01178779>.
- Rustichini, Aldo (2015). "The role of intelligence in economic decision making". In: *Current Opinion in Behavioral Sciences* 5. Neuroeconomics, pp. 32–36.
- Schelling, Thomas C. (1971). "Dynamic Models of Segregation". In: *Journal of Mathematical Sociology* 1, pp. 143–186.
- Schuck-Paim, Cynthia, Lorena Pompilio, and Alex Kacelnik (2004). "State-Dependent Decisions Cause Apparent Violations of Rationality in Animal Choice". In: *PLoS Biol* 2.12, pp. 2305–2315. ISSN: 15457885. DOI: [10.1371/journal.pbio.0020402](https://doi.org/10.1371/journal.pbio.0020402).
- Scotchmer, Suzanne (1991). "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law". In: *Journal of Economic Perspectives* 5.1, pp. 29–41. URL: <http://ideas.repec.org/a/aea/jecper/v5y1991i1p29-41.html>.

- Shafir, Sharoni, Tom A. Waite, and Brian H. Smith (2002). "Context-dependent violations of rational choice in honeybees (*Apis mellifera*) and gray jays (*Perisoreus canadensis*)". In: *Behavioral Ecology and Sociobiology* 51.2, pp. 180–187.
- Shchepetova, Anastasiya (2012). *Strategic limitation of price comparison by competing firms: An experimental study*. Working Paper. Toulouse School of Economics.
- Simonson, Itamar (2014). "Vices and Virtues of Misguided Replications: The Case of Asymmetric Dominance". In: *Journal of Marketing Research* 51.4, pp. 514–519. DOI: [10.1509/jmr.14.0093](https://doi.org/10.1509/jmr.14.0093).
- Slovic, Paul (1966). "Risk-Taking in Children: Age and Sex Differences". English. In: *Child Development* 37.1, pp. 169–176. ISSN: 00093920.
- Spiegler, Ran (2006). "Competition over agents with boundedly rational expectations". In: *Theoretical Economics* 1.2, pp. 207–231.
- (2011). *Bounded rationality and industrial organization*. Oxford University Press.
- Stiglitz, Joseph E (2007). "Economic foundations of intellectual property rights". In: *Duke LJ* 57, p. 1693.
- Synovate on behalf of the Central Office of Information, Food Standards Agency (2005). *Quantitative Evaluation of Alternative Food Signposting Concepts: Report of Findings*. Tech. rep. West Malling, UK: Synovate.
- Tajfel, Henri and John C Turner (1979). "An integrative theory of intergroup conflict". In: *The social psychology of intergroup relations* 33.47, p. 74.
- Tenbrunsel, Ann E. and David M. Messick (1999). "Sanctioning Systems, Decision Frames, and Cooperation". In: *Administrative Science Quarterly* 44.4, pp. 684–707.
- Teyssier, Sabrina (2012). "Inequity and risk aversion in sequential public good games". In: *Public Choice* 151 (1-2), pp. 91–119.
- Thaler, Richard H. and Cass R. Sunstein (2008). *Nudge*. Yale University Press.
- Tversky, Amos (1972). "Elimination By Aspects: A Theory Of Choice". In: *Psychological Review* 79.4, pp. 281–299. ISSN: 0033295X.
- Tversky, Amos and Daniel Kahneman (1992). "Cumulative prospect theory: An analysis of decision under uncertainty". In: *Journal of Risk and Uncertainty* 5.4, pp. 297–323.
- Tversky, Amos and Itamar Simonson (1993). "Context-Dependent Preferences". In: *Management Science* 39.10, pp. 1179–1189. ISSN: 0025-1909. DOI: [10.1287/mnsc.39.10.1179](https://doi.org/10.1287/mnsc.39.10.1179).
- Vyth, Ellis L et al. (2010). "Actual use of a front-of-pack nutrition logo in the supermarket: consumers' motives in food choice". In: *Public Health Nutrition* 13 (11), pp. 1882–1889. ISSN: 1475-2727. DOI: [10.1017/S1368980010000637](https://doi.org/10.1017/S1368980010000637).
- Vyth, Ellis L et al. (2012). "Methodological quality of front-of-pack labeling studies: a review plus identification of research challenges". In: *Nutrition reviews* 70, pp. 709–20.
- Wagner, Gert G., Joachim R. Frick, and Juergen Schupp (2007). "The German Socio-Economic Panel Study (SOEP): Scope, Evolution and Enhancements". In: *Schmollers Jahrbuch* 127 1.1, pp. 139–169.
- Wichmann, Felix A and N Jeremy Hill (2001). "The psychometric function: I. Fitting, sampling, and goodness of fit". In: *Perception & psychophysics* 63.8, pp. 1293–1313.
- Yamagishi, Toshio (1988). "Exit from the group as an individualistic solution to the free rider problem in the United States and Japan". In: *Journal of Experimental Social Psychology* 24, pp. 530–542. DOI: [10.1016/0022-1031\(88\)90051-0](https://doi.org/10.1016/0022-1031(88)90051-0). URL: [http://dx.doi.org/10.1016/0022-1031\(88\)90051-0](http://dx.doi.org/10.1016/0022-1031(88)90051-0).

Yang, Sybil and Michael Lynn (2014). "More Evidence Challenging the Robustness and Usefulness of the Attraction Effect". In: *Journal of Marketing Research* 51.4, pp. 508–513.

Appendices

Appendix A

Curriculum Vitæ

Paolo Crosetto - Italian citizen, June 20, 1979.
 Institut National pour la Recherche Agronomique
 Grenoble Applied Economics Laboratory
 1241 Rue des Résidences – Domaine Universitaire
 38400 Saint Martin d'Hères, France

Phone: +33.476.82.58.31
 Email: paolo.crosetto@inra.fr
 Homepage: <http://paolocrosetto.wordpress.com>

Current position

2013- Chargé de Recherche – INRA-GAEL National Institute for Agricultural Research – Grenoble, France
 Full-time researcher.

Education

- 2010 Ph.D., Economics, University of Milan.
 Department of Economics, Management and Quantitative Methods.
 Thesis advisors: Antonio Filippin and John D. Hey (York).
 Thesis in Experimental Economics. Experimental and Agent-Based essays on the economics of Open Source.
- 2003 Degree (Laurea), International Political Sciences, University of Turin, Italy.
 4-years degree (BA-MA), 110/110 *summa cum laude*. Thesis advisor: Carlo Boffito.
 Thesis in Economics of Transition. Case study: the Russian transition from plan to market.
- 2001 Erasmus Programme. Faculty of Economics, University of Nice-Sophia Antipolis (France)

Past Employment

- 2011-2013 Post-Doc at Max Planck Institute of Economics – Strategic Interaction Group – Jena, Germany.
 Full-time researcher
- 2010 Post-Doc at LUISS Guido Carli – Department of Economics and Finance – Rome, Italy.
 Research and teaching duties.

- 2005-2006 Research associate at Urge – Research Unit on European Governance – Turin, Italy.
Junior researcher
- 2005 Consultant at DASEIN – Turin-based public sector HR and continued education consultancy.
Consultant for several local authorities in North-western Italy. Teacher of IT skills.
- 2003-2005 Junior researcher at Cirpet – Research Centre on Emerging and Transition Countries – Turin, Italy.
Junior researcher. Coordinator of undergraduate research group.

Visiting and Temporary Positions

- 2011 (May) Invited visiting researcher, Department of Economics, University of York, UK.
- 2010 (November) Invited visiting researcher, Jena Graduate School “Human Behaviour in Social & Economic Change”, Jena, Germany.
- 2010 (October) Invited visiting researcher, Department of Economics, University of York, UK.

Invited Seminars

Total invited seminars: N = 23

- 2018 (forthcoming) Paris Economics and Psychology Seminar, Paris, France; (forthcoming) BETA seminar, Strasbourg, France; (forthcoming) Chair for microeconomics, Göttingen, Germany.
- 2017 (June) ALISS, INRA, Ivry sur Seine, France; (April) Scientific Flash presentation at the competition ‘Ma thèse en 180 secondes’ - Grenoble, France; (March) VALFREE Workshop, Rennes, France;
- 2016 (June) GK-EIC, University of Jena, Germany; (May) CREM, University of Rennes, France; (February) Department of Economics, Radboud University, Nijmegen, Netherlands
- 2015 (December) LAMETA, University of Montpellier, France ; (October) CEAR-CSAE Workshop “Eliciting Subjective Beliefs and Risk and Time Preferences in Developing Countries”, Oxford, UK; (May) ETH Zürich, Switzerland.; (April) Invited lecturer at AgroParisTech, Paris, France
- 2014 (December) University of Lausanne, Switzerland.; (November) Department of Economics and Related Studies, University of York, UK.; (April) CEGE, Georg-August-Universität Göttingen.
- 2013 (June) DIW, Berlin; (May) GAEL, Grenoble Applied Economics Laboratory, Grenoble.; (March) Centre d’Economie de la Sorbonne, University of Paris 1.
- 2012 (February) Department of Economics, University of Turin.
- 2011 (December) Department of Economics, Management and Quantitative Methods, University of Milan. ; (December) Faculty of Economics, Georg-August-Universität Göttingen.

2010 (November) Graduate College Economics of Innovative Change, Friedrich Schiller Universität Jena.

Conference Presentations

Total presentations: N = 35

- 2017 (October) 12th Nordic Conference, Gothenburg, Sweden; (August) EAAE 2017, Parma, Italy; (June) ASFEE 2017 - Rennes, France.
- 2016 (June) ASFEE 2016 - ESSEC Cergy, France; (April) IMEBESS - Rome, Italy; (December) Economics of Innovation - Grenoble, France.
- 2015 (October) Journées Jeunes Chercheurs INRA - Rennes, France ; (March) 143rd Joint EAAE-AAEA - Consumer behavior in a changing world, Naples, Italy; (April) NIBS 2015 Workshop, Nottingham, UK
- 2014 (November) Neuroeconomics and Food Workshop, Grenoble, France; (September) ESA Europe, Prague, Czech Republic; (August) EEA Conference, Toulouse, France; (May) ASFEE Conference, Besançon, France; (March) INRA Workshop "Farmers' Risk Preferences: Elicitation and Use", Rennes, France; (February) Ecole Chercheur Eco-Conception, INRA, Poitiers, France
- 2013 (November) MPI Autumn Workshop, Jena, Germany; (June) ASFEE conference 2013, Lyon, France; (May) BEElab conference 2013, Florence, Italy.; (April) IMEBE conference 2013, Madrid, Spain.; (February) ADRES Doctoral Conference, Strasbourg, France
- 2012 (September) 7th Nordic Conference on Behavioral and Experimental Economics, Bergen, Norway.; (September) ESA, Economic Science Association, European Meeting 2012, Cologne, Germany.; (June) IMPRS-ETH Workshop, Law and Economics of Intellectual Property, Schloss Ringberg, Germany ; (March) IMEBE conference 2012, Castellón, Spain.
- 2011 (July) ESA, Economic Science Association, European Meeting 2011, Luxembourg.; (July) ESA, Economic Science Association, World Meeting 2011, Chicago, IL, USA.; (June) Workshop Experimental Methods and Economic Modeling, Capua, Italy.
- 2010 (July) ESA, Economic Science Association, World Meeting 2010, Copenhagen, Denmark.; (June) FUR XIV, Foundations of Utility and Risk, Newcastle Upon Tyne, UK.; (June) OSS2010, 6th International Conference on Open Source Systems, Notre-Dame, IN, USA.; (April) VII LabSi Workshop on Experimental and Behavioural Economics, Siena, Italy
- 2009 (November) DIME International Workshop Experimental methods and Innovation, Max Planck Institute of Economics, Jena, Germany.; (July) 3rd FLOSS International Workshop on Free/Libre Open Source Software, Padova, Italy.; (May) Florence Behavioural and Experimental Economics Workshop, Florence, Italy.
- 2004 (April) UNECE Expert Meeting on Good Governance for SMEs, Geneva, Switzerland.

Teaching Experience

Year	Level	Role	Title	Language	Institution	Hours
2017-18	Doctorate	Teacher	Experimental Methods in Economics	English	UGA	12
	Master 2	Teacher	Experimental Economics: Energy and Environment	French	UGA	12
	Master 1	Teacher	Introduction to R – data wrangling & plotting	French	UGA	12
	Undergrad 1	Teacher	Introductory Microeconomics	English	GEM	45
2016-17	Master 2	Teacher	Experimental Economics: Energy and Environment	French	UGA	12
	Undergrad 2	Teacher	Introductory Microeconomics	English	GEM	30
	Undergrad 1	TA	Introductory Statistics	French	UGA	40
2015-16	Undergrad 1	TA	Microeconomics	French	UGA	12
	Undergrad 1	TA	Microeconomics for Industrial Engineering	French	INP	35
2014-15	Undergrad 1	TA	Microeconomics	French	UGA	12
	Undergrad 1	TA	Microeconomics for Industrial Engineering	French	INP	35
2012-13	Doctorate	Teacher	Group Assignment in Experimental Economics	English	IMPRS	12
2009-10	Undergrad 1	TA	Games and Strategy	English	LUISS	20
	Undergrad 1	TA	Microeconomics and Finance	English	LUISS	20
	Master 2	TA	Advanced Microeconomics	English	UniMi	20
2008-09	Master 2	TA	Advanced Microeconomics	English	UniMi	20
	Undergrad 1	TA	Microeconomics	Italian	UniMi	30
2004-05	Master 1	TA	Economics of Transition	English	UniTo	20
2003-04	Undergrad 3	TA	Economics of Transition	English	UniTo	20
Total class hours						419

Institutions cited in the table:

UGA Université Grenoble-Alpes, Grenoble, France.

GEM Grenoble Ecole de Management, Grenoble, France.

IMPRS International Max Planck School on Adapting Behavior in a Fundamentally Uncertain World, Jena, Germany.

INP Institut National Polytechnique de Grenoble, Grenoble, France.

LUISS Libera Università degli Studi Sociali Guido Carli, Rome, Italy.

UniMi Università degli Studi di Milano, Milan, Italy.

UniTo Università degli studi di Torino, Turin, Italy.

Refereeing, Editing and other Academic activities

Organizer First Winter Workshop: Behavioral and Experimental Economics of Food Consumption – Au-trans, France, January 25-27, 2017.

- Referee **N = 17.** Journal of the European Economic Association (1) – Journal of Economic Behavior and Organization (1) – Journal of Behavioral and Experimental Economics (2) – Experimental Economics (2) – Research Policy (1) – Nature Human Behavior (1) – Journal of Evolutionary Economics (1) – PLOS one (2) – Bulletin of Economic Research (3) – STAT (1) – Feminist Economics (1) – L'Actualité Economique (1)
- Advisor **N = 3.** GAČR – Czech Science Foundation (2011-2015); NSF – US National Science Foundation (2015); Comité de Thèse de Nicolas Pasquier (2016-2018); Oxford Nuffield Centre for Experimental Social Sciences ethical review Board
- Recruitment **N = 3.** University of Milan post-doc evaluation committees (2015); Member of the recruitment committee for Maitre de Conférences, Université Grenoble-Alpes (2016, 2017).
- Seminar In charge of GAEL external & internal seminars (with Adélaïde Fadhuile) (2016, 2017)
- Admin Member of the 'comité de pilotage' of the A2I project of functional renovation of the scientific platforms and laboratories in Grenoble INP – Site Viallet, Grenoble.

Languages (CEFR scale)

Italian: native speaker. **English:** C2. **French:** C2. **German:** B2 (but rusted). **Spanish:** B1.

Appendix B

Publications

Articles in Refereed Academic Journals [$N = 13$]

- 2017 P. Crosetto, A. Gaudeul. Choosing not to compete: Can firms maintain high prices by confusing consumers?
Journal of Economics Management and Strategy, forthcoming, CNRS 1.
- A. Gaudeul, P. Crosetto, G. Riener. Better stuck together or free to go? Of the stability of cooperation when individuals have outside options.
Journal of Economic Psychology, Volume 59, April 2017, Pages 99–112. , CNRS 2.
- P. Crosetto and A. Filippin. The Sound of Others: Surprising Evidence of Conformist Behavior.
Southern Economic Journal, Volume 83, Issue 4, April 2017, pages 1038–51, CNRS 2.
- P. Crosetto, A. Lacroix, L. Muller, B. Ruffieux. Modification des achats alimentaires en réponse à cinq logos nutritionnels
Cahiers de Nutrition et Diététique, Volume 52, Issue 3, June 2017, Pages 129–133
- 2016 A. Filippin and P. Crosetto. A Reconsideration of Gender Differences in Risk Attitudes.
Management Science, Volume 62, issue 11, November 2016, CNRS 1eg.
- J. Brüggemann, P. Crosetto, L. Meub and K. Bizer. Intellectual property rights hinder sequential innovation: experimental evidence.
Research Policy, Volume 45, Issue 10, December 2016, Pages 2054–2068, CNRS 1.
- P. Crosetto and A. Filippin. A Theoretical and Experimental Appraisal of Five Risk Elicitation Methods.
Experimental Economics, September 2016, Volume 19, Issue 3, pp 613–641, CNRS 1
- P. Crosetto, L. Muller and B. Ruffieux. Helping consumers with a front-of-pack label: numbers or colours? Experimental comparison between Guideline Daily Amount and Traffic Light in a diet-building exercise.
Journal of Economic Psychology, Volume 55, August 2016, Pages 30–50, CNRS 2.
- P. Crosetto and A. Gaudeul A Monetary Measure of the Strength and Robustness of the Attraction Effect
Economics Letters, Volume 149, December 2016, Pages 38–43, CNRS 3.

A. Filippin and P. Crosetto. Click'n'Roll: No Evidence of Illusion of Control
De Economist, September 2016, Volume 164, Issue 3, pp 281–295, CNRS 3

P. Crosetto and L. Muller and B. Ruffieux. Réponses des consommateurs à trois systèmes d'étiquetage nutritionnel face avant.
Cahiers de Nutrition et Diététique, June 2016, Volume 51, Issue 3, pp. 124-131

2015 P. Crosetto, A. Filippin and J. Heider. A Study of Outcome Reporting Bias Using Gender Differences in Risk Attitudes.
CESifo Economic Studies, Vol.61, Issue 1, pp.239-262.

2013 P. Crosetto and A. Filippin. The “Bomb” Risk Elicitation Task.
Journal of Risk and Uncertainty, Vol.47, Issue 1, August 2013, CNRS 2.

Refereed and Published Conference Proceedings [$N = 1$]

2010 P. Crosetto. To Patent or Not to Patent: A Pilot Experiment on Incentives to Copyright in a Sequential Innovation Setting.
Open Source Software: New Horizons. IFIP Advances in Information and Communication Technology. Volume 319, pp 53-72.

Other Publications - in Italian [$N = 2$]

2007 P. Crosetto. La Direttiva Bolkestein.
Biblioteca della Libertà, N. 186, January-March 2007.

2006 P. Crosetto. L'economia europea di fronte alla competizione globale: pronti per la ripresa o bloccati a metà strada?
Biblioteca della Libertà, N.184, July-September 2006.

Working Papers

R&R [$N = 1$]

Crosetto and T. Regner. Crowdfunding: Determinants of success and funding dynamics.
GAEL Working Paper 2015-02.

Submitted [$N = 4$]

P. Crosetto and A. Filippin. Safe options induce gender differences in risk attitudes.
GAEL Working Paper 2017-05

T. Regner and P. Crosetto. The experience matters: participation-related rewards increase the success chances of crowdfunding campaigns.

GAEL Working Paper 2017-04

P. Crosetto, O. Weisel and F. Winter. A flexible z-Tree implementation of the Social Value Orientation Slider Measure (Murphy et al. 2011) – Manual.

Jena Economic Research Papers 2012-062.

J. Bone, P. Crosetto, J. Hey and C. Pasca. Chance versus choice: eliciting attitudes to fair compensations.

Jena Economic Research Papers 2013-029.

Working Papers [N = 4]

P. Crosetto and M. Mantovani. Availability of Information and Representation Effects in the Centipede Game.

Jena Economic Research Paper 2012-051.

P. Crosetto, W. Güth, L. Mittone and M. Ploner. Motives of Sanctioning: Equity and Emotions in a Public Good Game with Punishment.

Jena Economic Research Papers 2012-046.

P. Crosetto. Turning private vices into collective virtues: a simple model and an experiment on the SourceForge development community.

University of Milan DEAS Working Paper 14-2009.

P. Crosetto. The American productivity miracle and Europe: the end of convergence?

Urge Working Paper 3/2005.