



HAL
open science

Choosing in a Large World: The Role of Focal Points as a Mindshaping Device

Lauren Larrouy, Guilhem Lecouteux

► **To cite this version:**

Lauren Larrouy, Guilhem Lecouteux. Choosing in a Large World: The Role of Focal Points as a Mindshaping Device. 2018. halshs-01923244

HAL Id: halshs-01923244

<https://shs.hal.science/halshs-01923244>

Preprint submitted on 15 Nov 2018

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.



UNIVERSITÉ
CÔTE D'AZUR



Université
Nice
Sophia Antipolis



CHOOSING IN A LARGE WORLD: THE ROLE OF FOCAL POINTS AS A MINDSHAPING DEVICE

Documents de travail GREDEG
GREDEG Working Papers Series

LAUREN LARROUY
GUILHEM LECOUTEUX

GREDEG WP No. 2018-29

<https://ideas.repec.org/s/gre/wpaper.html>

Les opinions exprimées dans la série des **Documents de travail GREDEG** sont celles des auteurs et ne reflètent pas nécessairement celles de l'institution. Les documents n'ont pas été soumis à un rapport formel et sont donc inclus dans cette série pour obtenir des commentaires et encourager la discussion. Les droits sur les documents appartiennent aux auteurs.

The views expressed in the GREDEG Working Paper Series are those of the author(s) and do not necessarily reflect those of the institution. The Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate. Copyright belongs to the author(s).

Choosing in a Large World: The Role of Focal Points as a Mindshaping Device

Lauren Larrouy and Guilhem Lecouteux

GREDEG Working Paper No. 2018-29

Abstract: The aim of this paper is to offer a theory of coordination that considers the role of the context within which the individuals interact, and to develop a rigorous analysis of salience and focal points. This requires dealing with how agents choose in ‘large worlds’ (in Savage’s sense). We highlight the role of mindshaping in the formation of individual preferences and beliefs and show how social focal points can generate prior beliefs. We conclude by discussing normative implications of our analysis, since it suggests that agents are socially-embedded entities, whose preferences and beliefs are shaped by social dynamics and norms.

Keywords: coordination, mindshaping, belief formation, preference formation, large world.

Subject classification codes: B41, C72, D81

Lauren Larrouy: *Université Côte d’Azur, CNRS, GREDEG, France*. Postal address: 250, rue Albert Einstein, 06560 Valbonne, France. Email: lauren.larrouy@gredeg.cnrs.fr

Guilhem Lecouteux (corresponding author): *Université Côte d’Azur, CNRS, GREDEG, France*. Postal address: 250, rue Albert Einstein, 06560 Valbonne, France. Email: guilhem.lecouteux@unice.fr

Humans routinely coordinate on complex activities, whether it be playing a matching game in a lab experiment, driving, working within a team, co-authoring an academic paper, or meeting in a crowded place such as a train station or during a concert. An explanation of this ability to coordinate with others is that we share a capacity to correctly represent each other's mental states and beliefs, and can thus accurately anticipate the behaviours of others. Different theories of *mindreading* – the capacity to ‘read’ each other's minds – have been suggested in the literature,ⁱ though they systematically start from the principle that the key to human coordination is our ability to form complex epistemic states about each other's mental states, especially beliefs and desires (Zawidzki 2018). This is also the approach implicitly endorsed in the epistemic program in game theory, according to which the analysis of strategic behaviours can be reduced to Bayesian decision theory, while taking hierarchies of beliefs as an input of decision-making.

There are however some methodological issues with the project of developing rigorous Bayesian foundations for game theory.ⁱⁱ We can in particular question the usefulness of an approach that requires players holding complex beliefs about each other's actions and beliefs, without explaining *how* players could form such beliefs. The lack of a theory of belief formation implies that Bayesian game theory – just as classical game theory – is unable to explain successful coordination in simple matching games. Although mindreading may offer a justification for the formation of common belief of events and of the rationality of the players (Larrouy and Lecouteux 2017, proposition 1), it remains insufficient to determine unambiguously to which beliefs the players will *in fine* converge. The main issue of Bayesian game theory is indeed that the beliefs that the players are supposed to use as an *input* of their reasoning, are by construction the beliefs at the equilibrium, i.e. the *output* of their reasoning (Lecouteux 2018a).

The usual strategy used by game theorists to solve coordination puzzles is to suppose the existence of *focal points*, and to suppose that people are more likely to believe that the focal point is the ‘correct’ choice. Apart from a few attempts to develop a rigorous analysis of salience and focal points (Bacharach 1993, Sugden 1995, Casajus 2000, Janssen 2001), those notions remain loosely defined and merely offer an *ad hoc* explanation. The conceptual issue here is that game theory is ‘an internally closed procedure which operates according to fixed rules’ (von Neumann 1983 [1931], 61-62, quoted in Giocoli 2003, 15), and therefore that all the relevant information for the players *must be* included in the mathematical description of the game. Developing an empirically

robust theory of games would however require including the processes by which the players use the ‘background’ of the game to choose their strategy, i.e. would require including the role of the *context* in game theoretical analysis.

An alternative explanation for human coordination that has been highlighted by cognitive scientists is the process of *mindshaping*, according to which it is the ability to *shape* each other’s minds that is the key to our evolutionary success. It is defined by Zawidski (2013, p. xiii) as ‘a group accomplishment, involving simultaneously interpretive and regulative frameworks that function to shape minds.’ It includes social devices such as norms, conventions, or institutions, which regulate individual behaviours and social practices. The aim of such devices is to generate a ‘cognitive homogenisation’ (Zawidski, 2013, p.65), so that people’s behaviours become easily interpretable. Successful coordination is not the result of a highly intellectualised process that requires people to solve the intractable problem of correctly inferring the independently mental states of others. It is rather because we share some common perceptions *ex ante*, and that we have been exposed to the same norms and conventions, that we can easily anticipate each other’s behaviours, based on our own experiences. Mindshaping – by shaping our minds in similar ways – is a prerequisite to successful coordination, since it guarantees that our own mental states (whether they be experiences, perceptions, reasoning mode or beliefs), are likely to be similar to the mental states of the other individuals with whom we interact.

The aim of this paper is to develop a theory of coordination that considers the role of the *context* within which the individuals interact. In particular, we will emphasise the role of mindshaping in the formation of a common intersubjective background among the players, allowing *in fine* players to identify and coordinate on the same focal points. The explicit introduction of the processes by which the players represent real and complex decision problems as simpler, more tractable, ones, however requires tackling the problem set aside by Savage (1954) of developing a theory of choice in *large* worlds. We will start from Savage’s discussion of small and large worlds to introduce a formal definition of salience and focal points, by relying conceptually on Bacharach’s variable frame theory, and formally on Mandler *et al*’s (2012) model of choice by checklist. We will argue that the prior beliefs of the players can be derived from their representation of

the common perception of the game, and that this shared perception is the outcome of mindshaping through evolutionary dynamics. Once the shared representation of the game and the prior beliefs of the players have been identified, the standard tools of Bayesian game theory can be applied.

We begin with two illustrations (a standard matching game in Paris, and the Brexit negotiations) to highlight the main features of our analysis of ‘open’ systems (section 1). We then recall Savage’s definition of small and large worlds, and highlight the role of focal points as a factor of cognitive homogenisation, allowing the players to adopt the same representations of a large world (section 2). We then model the formation of the beliefs of Bayes rational agents in a large world, and discuss the role of focal points in the evolutionary dynamics of mindshaping processes (section 3). We conclude by discussing normative implications of our framework, and more specifically the too restrictive understanding of the notion of ‘context’ in behavioural welfare economics (section 4).

1. Two illustrations

1.1 Meeting in Paris

Suppose that you are asked to meet someone (about whom you have no additional information for the moment – she could be a tourist, a colleague, a friend, etc.) in Paris in exactly a week from now at noon. Where will you wait for the other? This is a standard ‘matching game’, in which you intend to go to the same location than the other individual, whatever this place is.

Probably one of the first meeting points – if you are not a Parisian – that spontaneously came to your mind was the Eiffel Tower, which is the most iconic place of Paris. Note that you can confidently feel that we would *all* recognise the Eiffel Tower as a focal point for a meeting in Paris (including the other individual you have to meet). This is because, as part of the community of ‘non-Parisians’ (with a limited knowledge of the city), it seems reasonable to assume that we would all feel the same than how you felt when thinking about a focal point in Paris. The fact that you identified the Eiffel Tower as focal may give some evidence about the focal point that is identified by the other members of the community of non-Parisians.

Suppose now that you live and work in Paris: like most Parisians, the Eiffel Tower is not particularly salient for you anymore, and a more natural meeting point would typically be the station Châtelet Les Halles, the main commuter train hub in Paris. From the perspective of a Parisian, this station stands as a more obvious solution to the game than the Eiffel Tower. However, as many Parisians, you also probably have a favourite café for meeting your friends after work, another for having brunch or for drinking your morning café before going to work, a favourite bistro, etc. Depending on the context (whether you have to meet the other in the morning, at noon, after work, during a weekend, etc.) you may spontaneously identify different places as the most salient ones to meet the other individual.

Given that you must meet the other individual at noon, this particular context may trigger the idea that you should meet in the restaurant you usually go for lunch (for instance the one close to your workplace if it is a weekday). This specific restaurant is then of ‘primary salience’ for you (Metha *et al.* 1994a,b; Bacharach and Bernasconi, 1997), it is the obvious meeting point that spontaneously comes to your mind. We will define later in the paper the primary salient option as an *individual focal point*.

However, without any additional information about the other, there is little chance that she will also identify your favourite restaurant as a focal point. This means that you have to anticipate how the other will frame the game, and what focal point she is likely to identify. The element of the context that seemed important for you (it is lunchtime, so a restaurant is a good meeting point) may not be the element that the other will consider as important. This means that you have to ‘think about salience strategically’ (Bacharach and Bernasconi, 1997, 39). If it turns out that the other individual is one of your colleagues with whom you regularly go to this restaurant for lunch, then it seems reasonable to think that going to the restaurant is primary salient for her too. But if you simply learn that the other individual is a Parisian (without any additional information), then you have to think about what is of primary salient for a ‘Parisian’: here Châtelet Les Halles seems to be a good candidate. By supposing that Châtelet Les Halles is of primary salience for the other (as a Parisian), it becomes of secondary salience for you (Bacharach and Bernasconi, 1997, 38). We will define later such secondary salient option as a *subjective social focal point*.

We could continue the illustrations with many other cases. For instance, if you have no information about the other (you are thus in a complete uncertainty) – she could

either be the colleague with whom you regularly go for lunch or a foreign tourist – it seems safer to restrict the set of possible meeting points to what is shared by virtually everybody, i.e. the Eiffel Tower. If you both identified the same meeting point as a subjective social focal point (i.e. the Eiffel Tower with a tourist, Chatelet les Halles with a Parisian, the restaurant with your colleague), then we will call such point an *objective social focal point*.

The existence of objective social focal points (such as the Eiffel Tower as a default option, Chatelet les Halles for Parisians, the restaurant for your colleagues) is the product of mindshaping, *via* routines and recurrent interactions with others. It is because many Parisians meet at Châtelet that it became focal for them, and because you usually go with your colleagues to the same restaurant that it is also focal for your smaller community. The initial problem, which was of an infinite complexity (they are uncountable meeting points in Paris), became almost trivial simply thanks to the fact that both you and the other individual belong to a same community ('everybody', 'Parisian', or 'colleagues'), with a set of shared perceptions and norms. As emphasised by Schelling (1980[1960] pp. 57-58):

'A prime characteristic of most of these "solutions" to the problems, that is, of the clues or coordinators or focal points, is some kind of prominence or conspicuousness. But it is a prominence that depend on time and place and who the people are.'

A problem that remains to be solved, however, is the question of the rationalisation of focal point play: although it seems 'obvious' that a focal point is a perfect guide to coordination, we still lack Bayesian foundations for such type of play. One of the objective of the paper – in addition to offering an approach to model how players frame the initial problem into a simpler one – will be to argue that the most salient option in the game (as we think is commonly perceived by the members of the community) can be integrated as our 'gut feeling' regarding the outcome of the game, which constitutes a basis for the prior belief we will use when solving the game.

1.2 Brexit negotiations

We now turn to a slightly different situation, to illustrate how players may frame the initial problem into a much more tractable one. The idea is here to show that – unlike most of game theory that starts with already abstract situations (like meeting a random individual in Paris in a week from now) – the processes we intend to model also govern ‘real’ interactions. It will also give us the opportunity to introduce the main ingredients of our formal model to translate ‘large worlds’ into ‘small worlds’.

Consider the strategic interaction between the UK government and the European Union about the nature of their future relationship after the Brexit. The precise and exhaustive description of the set of actions (i.e. the details of the possible agreements between the EU, the UK, and third parties, the timing of the negotiation and the possible impacts of national elections, etc.) for both parties is far from being cognitively possible. A way of framing the problem for the two parties is to consider a restricted set of acts, and to focus only on very few features of the consequences of those acts. The UK may for instance consider that only two ‘properties’ of the possible outcomes are relevant from its perspective: the UK economic performance and migration flows (as this were the two central themes during the 2016 referendum campaign). The UK may also believe that the EU will evaluate the outcomes mainly according to one property, ‘economic performance of the EU’. In the game imagined by the UK, the set of possible outcomes is thus framed according to those combinations of properties. This will also determine how the UK perceives the available strategies for both players: the UK will propose an agreement to the EU, and the EU may accept or reject it (the ‘no-deal’ scenario is detrimental for the economic performance of both players, but is a good outcome in terms of migration flows for the UK). Among the set of agreements that the UK could offer (its set of strategies), the key elements will concern the two relevant properties for the UK, i.e. protecting the free trade between the UK and the EU (with a good outcome in terms of economic performance) while limiting the migration from the EU to the UK. The UK can represent its available strategies as (i) offer a free trade area, (ii) stay in the Single Market, or (iii) no offer (with a no deal scenario). Assuming that the EU only cares about its economic performance, it is in the EU interest to accept any offer, and to avoid a no-deal scenario. The UK can thus confidently offer only a free trade area, and wait for the positive reply from the EU – this was broadly the rhetoric of the ‘Leave’ campaign during the

referendum, because it was argued that it would be in the interest of the EU to stay in good terms with the UK after the Brexit.

The negotiations (from the UK's perspective) are thus framed as a game in which they can choose between three strategies (free trade area, single market, no offer) and the EU between two strategies (accept or reject the offer). Since rejecting the offer seems to be a weakly dominated strategy for the EU, the best strategy for the UK is to offer a free trade area and to limit migration flows. If, however, the EU also evaluates the outcome according to the property 'deterrence' (on future exits from the EU), then a scenario of 'no-deal' may be acceptable by the EU, since an economic failure following Brexit could dissuade other countries from leaving the EU. Depending on which property (economic performance of the EU, or deterrence) the EU most favours, the evaluation of the outcomes may be radically different from what is framed in the UK's game representation, in which only the economic performance of the EU is considered by the EU.

Our framework aims at capturing the main features discussed here. Each player will frame first her set of strategies (based on the 'salience' of some acts). They will then try to frame the set of strategies of the other player, by a process of *simulation* (by imagining how they would represent the set of strategies, if they were in the other's shoes). Given the definition of the set of strategies for each player, each strategy profile will be assessed on the basis of various 'properties', over which the agents will have a preference relation (e.g. the EU may prefer the property 'deterrence' over 'EU economic performance'). We will model this preference over properties thanks to Mandler *et al* (2012) model of choice by checklist.ⁱⁱⁱ Given their set of strategies, and the evaluation of the outcomes, we obtain a game representation (a *small world*, in Savage's terminology) in which the only uncertainty concerns the strategy of the other player.^{iv} We will then assume that the players form their beliefs about the choice of the other by simulating their reasoning, as in Larrouy and Lecouteux (2017).

2. Games as large worlds

2.1 Open and closed problems

A game in normal form is a mathematical object defined by a triplet $G = \langle N, X, \Pi \rangle$, which defines sets of players, strategies, and payoffs. The reason why matching games constitute a theoretical puzzle is that the way we ‘write down’ the game as a payoff matrix usually conveys more information than what is included in the mathematical object (e.g. we label strategies with letters ‘A’ and ‘B’). A relatively faithful and neutral description of the mathematical object corresponding to a matching game would however be something like: ‘you have the same set of two strategies; if you both choose the same strategy you gain more than if you don’t; which strategy do you choose?’. The problem formulated in this way is genuinely puzzling because we simply would like to answer ‘the same strategy than the other’, but there is nothing in the description of the problem that allows us to differentiate in any way those two strategies. When considering decision problems faced by actual players, the game is necessarily embedded in a particular context (a lab experiment, a specific social setting), and – while the game as a mathematical object aims at capturing the strategic features of the interaction only – the context will provide clues to the individual about the choice to be made (i.e. drive on the left side of the road in England, but on the right side in France). The role of the ‘context’ in decision making is extremely intuitive – both as a way to intentionally discriminate between alternatives in matching games, but also as an unintentional factor that impacts our decisions (see e.g. Kahneman and Tversky (2000) on framing effects) – but remains a fundamentally vague theoretical object, as the background in which the game takes place is necessarily *external* to the game. Referring to such external factors (such as norms of language in the description of the game in Bacharach (1993) ‘variable universe games’) implies that the choice problem is ‘open’ rather than ‘closed’ or ‘completable’ (Binmore 1988, 1993).

A crucial issue of open choice problems is that we cannot *a priori* rely on the tools of subjective expected utility theory. Traditionally, expected utility theory is founded on an axiomatic derivation of probabilities and utilities from a set of observed preferences over acts (e.g. Ramsey 1926, Von Neumann and Morgenstern 1953, Savage 1954, Anscombe and Aumann 1963, Aumann and Drèze 2009). Those systems of axioms are

designed such that they lead to a unique class of state-independent utility functions and probability distributions, which rank the different acts according to their expected utilities. The primitive of analysis is the *choice* of the players, and the utility is defined *ex post*, as a representation of their behaviour (preferences are then called ‘behaviouristic’).

The approach to subjective expected utility in this paper significantly differs from this traditional approach, as we do not treat choices but *payoff* as the primitive of analysis. To avoid any confusion with the ‘utility’ defined *a posteriori* from individual preferences over acts, we will only use the term ‘payoff’, defined as ‘normalised measures of the values of the relevant outcomes to [the player], *in terms of her own interests, as judged by her*’ (Sugden 2015, p.144, emphasis in original). Unlike utility, which merely describes the preferences of the player, the payoff is defined as the *cause* of the preferences of the individual – I prefer the act *f* over the act *g*, because the associated expected payoff is higher^v (while the expected utility of *f* is higher than *g*, *because I prefer f*).

Taking payoffs rather than choices as the primitive of analysis means that we must investigate how players choose for given payoffs, i.e. how they form their beliefs and intentions. This requires a significant departure from the usual Bayesian analysis of strategic interactions, since this kind of problem cannot be considered as a ‘small world’, which is the only situation – as argued by Savage (1954), and more recently by Binmore (2009) – to which Bayesian decision theory could be applied:

[t]he models constructed by game theorists are small worlds almost by definition. So we can use Bayesian decision theory without fear of being haunted by Savage's ghost telling us that it is ridiculous to use his theory in a large world. Bayesian decision theory only ceases to be appropriate when attempts are made to include the minds of the players in the model to be considered' (Binmore 2009, pp.134-135)

If we endorse a mentalistic rather than behaviouristic interpretation of payoffs and beliefs^{vi} and intend to model the formation of players’ beliefs in a non-tautological way, then we must propose a framework for choosing in ‘large worlds’. One of the main implications of a mentalistic approach to Bayesian decision theory is that we need a

psychological theory of the formation of the players' prior beliefs – they are indeed defined *ex post* in a behaviouristic approach, together with the players' utility functions. The problem of choosing in a universe that is not completable has been little studied so far, and we will follow Binmore's suggestion who sees:

... one of the major purposes of studying foundational questions [in game theory] as being that of finding appropriate ways of closing the universe of discourse so as to *legitimize* the use of Bayesian methods in analysing particular games. (Binmore 1993, 322)

The approach we choose to close the universe will conceptually rely on Bacharach's variable frame theory, though the formalism will be slightly different to better catch Savage's distinction between small and large worlds. In a nutshell, we will consider that some features of the decision problem will be more *salient* than others, and the players will be 'guided' by those salient features to frame the initial large world as a tractable problem – a small world – for which Bayesian methods can be applied. An unconscious part of the process of framing will be driven by mindshaping, while we will also allow players to strategically frame their problems to facilitate coordination (as in Schelling 1960). In other words, the unconscious framing will offer a basis for the 'gut feelings' of the players when thinking about the outcome of the game (corresponding to the first solution that spontaneously comes to your mind, such as the Eiffel Tower), which should then be possibly revised to take into account possible inconsistencies (as in Larrouy and Lecouteux (2017), who characterise as 'massaged belief hierarchies' the nature of the beliefs that would survive this consistency check).

2.2 Small worlds, large worlds, and the grand world

We slightly adapt Savage's notations to define large and small worlds. We consider a set $S = \{s, s', \dots\}$ of *states of the world*, a set $\mathcal{C} = \{c, c', \dots\}$ of *consequences*, and define an *act* as an arbitrary function $\mathbf{A}: S \mapsto \mathcal{C}$ that assigns consequences to states of the world. We denote by \mathcal{A} the set of acts. A *world* is a triplet $W = \langle S_W, \mathcal{C}_W, \mathcal{A}_W \rangle$ that describes the set of possible acts \mathcal{A}_W and their consequences in \mathcal{C}_W , given the possible states of the world in \mathcal{S}_W .

A world W is *smaller* than a world W' if and only if the set of states of W is a partition of the set of states W' (W' is then a *larger* world than W). For instance, if the states of W describe the range of possible temperatures in Paris at noon – which may go from -25°C to 40°C – a smaller world W' can for instance be composed of the two states ‘below 0°C ’ and ‘above 0°C ’. The two states below and above freezing indeed form a partition of the range of possible temperature in Paris at noon.

Each element $s \in S_W$ offers a more or less precise description of the state about which the decision maker is uncertain. 16°C is for instance more precise than ‘above 0°C ’, though it could have been even more precise if the state described the distribution of temperature at the different weather stations in Paris. Note that we can almost always add arbitrarily detailed elements to the description of the state of the world (e.g. describe the temperature in Paris by the distribution of temperatures in Celsius degree, precise up to the third decimal, for each square meter of the city of Paris). But if my initial problem is whether I should wear a coat or not, I do not need such a precise description – it would be enough to partition the states of the world with a few states such that ‘below 15°C ’ and ‘above 15°C ’. 15°C seems indeed to be a reasonable threshold for many people to choose whether to wear a coat or not (but different people could consider that the relevant threshold – and then, the framing of the initial problem as a small world involving only two states – is lower or higher).

Developing a theory of choice in large worlds requires modelling how the agents frame the initial problem as a small world, with a ‘reasonable’ number of states of the world. We do not need to define a threshold in terms of cardinality of the set of states of the world to characterise a ‘small’ or ‘large’ world. The distinction is mostly qualitative, i.e. the differences between states of the world are meaningful for the choice under consideration.

Savage uses the two following proverbs to explain the nature of the difference between small and large worlds:

‘The point of view under discussion may be symbolized by the proverb, “Look before you leap,”, and the one which it is opposed by the proverb, “You can cross that bridge when you come to it.” When two proverbs conflict in this way, it is proverbially true that there is some truth in both of them, but rarely, if ever, can their common truth be captured by a single pat proverb. One must indeed look

before he leaps, in so far as the looking is not unreasonably time-consuming and otherwise expensive; but there are innumerable bridges one cannot afford to cross, unless he happens to come to them.' (Savage 1954, p.16)

A choice problem in which 'Look before you leap' is a reasonable principle of choice, i.e. in which the individual can anticipate all the possible consequences and plan in advance all her future moves given the possible states of the world, 'in so far as the looking is not unreasonably time-consuming and otherwise expensive' (Savage 1954, 16) is called a small world. Otherwise:

'Carried to its logical extreme, the "Look before you leap" principle demands that one envisage every conceivable policy for the government of his whole life (at least from now on) in its most minute details, in the light of the vast number of unknown states of the world, and decide here and now on one policy. This is utterly ridiculous, not – as some might think – because there might be later cause for regret, if things did not turn out as had been anticipated, but because the task implied in making such a decision is not even remotely resembled by human possibility' (Savage 1954, 16)

Our interpretation of Savage is that the distinction between small and large worlds is not of an ontological nature.^{vii} Although it might seem there exists a fundamental difference between being uncertain about 'Whether a particular egg is rotten' and 'The exact and entire past, present, and future history of the universe, understood in any sense, however wide' (Savage 1954, p.8), Savage sets the limit between small and large worlds with a criterion of *bounded cognitive ability*. A game of chess – although it is logically solvable by exploring all the possible sequences of moves – is for instance considered as a large world. Rather than anticipating all the future moves from a given position, a chess player will typically try to reach a position that seems good (a 'dream position'), and will start to think about her next moves only when arriving in that position (problems are thus treated sequentially). Similarly, in the case of your meeting in Paris, although the Eiffel Tower is a focal point, you will quickly realise that it is not clear where you should meet once arrived there. Probably you will then look for a meeting point close to the entrance, or try to find a spot where you can easily be seen and recognised. But when asked to meet

someone in Paris a few pages ago, you did not consider all those practical issues: you reframed the initial problem as a succession of simpler problems (first, find a good spot in Paris; second, find a good spot at the Eiffel tower).

Savage's sketch of a theory of large worlds (pp.82-91) starts from the 'grand situation', and considers that the agent will successively solve tractable problems by focusing on 'isolated decision situations'. Our framework starts from a similar principle, and we define the *grand world* $GW = \langle \mathcal{S}_{GW}, \mathcal{C}_{GW}, \mathcal{A}_{GW} \rangle$ as the 'largest' possible world. GW offers the most complete and detailed description of all the states of the worlds, acts, and of their respective consequences. Describing GW would require a perfect omniscience, so we must keep in mind that we use GW only as a theoretical tool to build a tractable model of framing in large worlds, and not as an actual world for which the agents could distinguish all the different states. In the case of the coordination problem in Paris, a GW-act could be something like 'go to a precise GPS coordinate, at a precise altitude'. This means that the act 'go to the Eiffel tower' corresponds to a significant subset of the GW-acts (the world in which this act is available is thus smaller than GW).

According to Savage, 'to cross one's bridges when one comes to them means to attack relatively simple problems of decision by artificially confining attention to so small a world that the "Look before you leap" principle can be applied there' (Savage 1954, p.16). He however confesses being 'unable to formulate criteria for selecting these small worlds', while believing 'that their selection may be a matter of judgment and experience about which it is impossible to enunciate complete and sharply defined general principles' (p.16). Bacharach's (1993, 2006) *variable frame theory* offers a framework explaining how players frame open problems as closed ones:

[I]n order to explain how someone acts, we have to take account of the representation or model of her situation that she is using as she thinks what to do. This model varies with the cognitive frame in which she does her thinking. Her frame stands to her thoughts as a set of axes does to a graph; it circumscribes the thoughts that are logically possible for her (not ever but at the time). In a decision problem, everything is up for framing. The preferences on which she acts, her alternatives ... So far from finding herself with given preferences over outcomes,

as traditional theory holds self-evident, these preferences depend upon the evaluative concepts that are uppermost in her mind. (Bacharach, 2006, 69)

Our approach will be conceptually similar to Bacharach's, though we will adopt a slightly different formalism. The main idea is that certain small worlds within the large world are more salient than others, i.e. that some representations 'seem' more natural from the perspective of the players compared to others. Chess players for instance do not consider exhaustively all the possible sequence of moves, because they can identify *ex ante* only a few patterns which seem more relevant than others (and good chess players will be able to recognise intuitively only the best available candidates moves). This means that the solutions that are 'intuitive' and 'uppermost in [the player's] mind' are of major importance for the individual's decisions, and constitute what we will call *individual* focal points.

2.3 Mindshaping and focal points

Schelling (1980[1960]) defines focal points as everything that is salient: a pattern of behaviour, a strategy, an outcome, etc. Anything that leads the players to perceive an outcome as the solution of the game is a focal point. A focal point therefore serves as a guide to coordination: it first *allows* coordination by making individual beliefs converge to a specific solution (Schelling 1980[1960] p. 56), and second *maintains* coordination by reinforcing the coherence of individual beliefs regarding everybody else's behaviour (Hédoin 2014, p. 366). Focal points can be used both as an input and as an output of successful coordination. Seen as a regulatory social device,^{viii} a focal point generates symmetric reasoning, and may induce stable patterns of behaviours.

The existence of focal points is part of the processes of mindshaping to the extent that it eases the cognitive homogenisation of the environment, both by aligning players' perceptions of the environment, and by reinforcing those perceptions thanks to the success of past coordination. If Châtelet les Halles is focal among the community of Parisians, it aligns the beliefs of Parisians about what is a good meeting point, and the fact that Parisians successfully coordinated by going to Châtelet les Halles justifies that it constitutes a good meeting point for Parisians. Even though my individual focal point can be to go to the restaurant, my subjective social focal point (what I believe I share with

the other individual) to go to Châtelet les Halles is reinforced if it turns out to be an objective social focal point (i.e. if all the Parisians consider it as a subjective social focal point).^{ix}

The cognitive processes allowing a successful coordination is a combination of *simulation thinking* and mindshaping. Simulation thinking, as conceptualised by Goldman (2006), consists in using one's own mind to simulate the reasoning of others. People anticipate the beliefs and behaviours of others by projecting their own perceptions, beliefs, and mode of reasoning in the mind of others – first by attributing their own beliefs in presence of uncertainty (e.g. I believe that the Eiffel Tower is focal for a foreign tourist, although I have no specific information about her), and second by using their own reasoning as a simulator of the reasoning of others (e.g. if I think it is 'rational' to go to the focal point, I will believe that you think it is 'rational' to go to the focal point too). This means that, in case of a full uncertainty about the other, the individual tends to attribute to the other her own beliefs about what she considers as focal.^x If I have no reason to believe that you could be unaware that the restaurant is a good meeting point, then my own individual focal point becomes a subjective social focal point.

It now clearly appears that simulation thinking is accurate and brings correct anticipation only if our minds have been shaped by the same models – and that it is indeed realistic to identify the Eiffel tower as focal, and that my mode of reasoning (e.g. expected payoff maximisation) is similar to yours. The role of mindshaping becomes decisive: it is only if we share a common and homogeneous perception of the environment that simulation may be effective – and that the points we identify as subjective social focal points are *objective* social focal points. If we have been exposed to the same set of social devices, such as conventions or institutions, collective narrative, etc, we will observe similarities among our minds so that we will tend to focus on the same features of reality (Zawidski, 2013, p. 86). Furthermore,

from the mindshaping perspective ... human social cognition relies centrally on our capacities and dispositions to shape each other and ourselves to conform to normative expectations that prevail in groups of likely interactants. (Zawidski 2013, p. 236)

An important point here is that my prior belief about how the other will behave depends on what I think we share in common: apart from the case of full uncertainty (where I tend to attribute my own individual focal point to the other), any information about the other – i.e. you are a Parisian, a colleague, or a tourist – will determine my belief about what we have in common, and therefore what we are likely to both identify as a social focal point.

Given the initial (and possibly infinitely complex) problem of meeting in Paris, mindshaping helps to ‘drastically reduce, and hence make manageable, the space of interactions in which we engage: we restrict the games we can play with each other ... and coordination is dramatically facilitated.’ (Zawidski, 2013, p. 58).^{xi} In other words, mindshaping reduces the set of properties that the players can have in common. It first determines our own small world representation (through the different properties that spontaneously come to our minds – we for instance does not think in terms of GPS coordinates, but of larger locations), and then our beliefs about the *shared* small world representation, depending on whom we are interacting with.

The role of mindreading (through the formation of complex epistemic states about others’ beliefs, intentions, or mode of reasoning) appears to be considerably reduced. It is indeed ‘only with prior, sophisticated mindshaping, ensuring cognitive homogeneity in populations of likely interactants, can distinctively human mindreading be reliable.’ (Zawidski, 2013, p. xvi). Mindshaping provides a ground for an accurate *fast and frugal* procedure of mindreading in simulation:

We are able to interpret each other’s behavior accurately and rapidly because most of our interactants are products of the same cultural mind-shaping influences as ourselves (Apperly, 2011, pp. 29, 160). We tend to interact with people who have been subject to the same kinds of pedagogy, who have tried to imitate the same kinds of role models, and who seek to conform to the same norms as ourselves. For this reason, our fast and frugal interpretations are typically accurate: the cognitive homogeneity on which they depend is a product of sociocultural mindshaping practices.” (Zawidski, 2013, p. 81)

Even if we do not share the exact same cultural models and cannot reach a perfect homogeneity, a partial overlapping of those models is in many cases sufficient to bring enough homogenization. And if ever the small world representation that the players have built turns out to be erroneous, they will build another small world until congruence is reached. An objective social focal point – by signalling the convergence of small-world representations – reinforces the beliefs and representations of the players, while it is initially the outcome of an evolutionary process of trials and errors.

3. Choosing in large worlds

The analytical framework we develop in this section should (i) explain how the players translate a large world into a small world, (ii) how they tend to identify the same focal points through the cognitive homogenisation of the environment, and (iii) characterise the beliefs that the agents may form as a way to rationalise the choice of focal points.^{xii} We will deal with the first part by considering that each individual is *aware* of a certain number of ‘properties’ of the outcomes, and have a ranking of those properties based on their salience. We will show that those two ingredients are sufficient to ensure that players choosing the most salient outcome choose as if they were maximising a strictly increasing payoff function. We then suppose that the individuals are able to distinguish between their own perception of the game (and the individual focal point they identify), and what they think is the shared perception of the game. An important addition to standard Bayesian game theory is that we next assume that this shared perception offers a basis for defining the prior beliefs of the players. We then show that, given this prior belief, it is indeed individually rational for the player to play their part of what they identify as a social focal point.

3.1 Preliminaries

We define GW as the grand world associated to a strategic interaction involving a set $N = \{1; \dots; n\}$ of players. We formalise how a player $i \in N$ will represent GW as a small world SG_i (note that we index by i because it is not sure that i 's representation of the grand world will be the same than the other players – cf. the case of Brexit above).

In the grand world GW , each combination of GW -acts determines a possible outcome of the strategic interaction (e.g. each of us stands in a precise GPS coordinate, at a precise altitude in Paris). The exhaustive set of outcomes is assessed unconsciously according to a set of *properties* P . Each property assigns a value to a combination of GW -acts: this can either be 0 or 1 (i.e. the property is satisfied vs. it is not), or a ranking with different values. We can thus partition the set of combinations of GW -acts, depending on the properties that are satisfied (or more generally, depending on the index associated).

In the case of the meeting in Paris, a possible property is ‘well-known touristic places in Paris’, which ranks e.g. the Eiffel tower first, then Notre Dame, the Louvres museum, etc.; another could be ‘nice places to have lunch’, with the ranking of your favourite restaurants. Note that, in the ‘linguistic’ definition of the two properties we just introduced, it is likely that two individuals may consider different rankings for e.g. touristic places (e.g. rank Montmartre before Notre-Dame). To avoid such confusions, we will define a property analytically as a function $P(j): \mathcal{A}_i \times S \mapsto \mathbb{N}$. A property $P(j)$ associates an index from the set of natural integers^{xiii} to each GW -act for a player i , given a possible state of the world (i.e. the action of the other players) – equivalently, this corresponds to a vector of GW -acts.

Each player is aware of a certain number of properties, that we will call her *awareness set* $\mathcal{P}_i = \{P(j)\}_{j \in J_i}$ with $J_i \subseteq \mathbb{N}$ a set of indices (which can be finite or not). From the perspective of player i , it is thus possible to partition the set of outcomes depending on the properties satisfied. If a combination of GW -acts does not satisfy any property, then the player is not aware of this specific outcome (e.g. if you are not an economist, you will probably not be aware that the campus of Paris School of Economics is a possible meeting point).

The two basic ingredients of our framework are: (i) the awareness set, $\mathcal{P}_i = \{P(j)\}_{j \in J_i}$, and (ii) a complete and transitive relation \succsim_i over this awareness set. The resulting ordering of properties is called the *checklist* of player i and is denoted $\mathcal{C}_i = \{P(j)\}_{\succsim_i}$. The term of ‘checklist’ comes from Mandler *et al* (2012), who study the behaviour of individuals who choose by following a checklist (i.e. I first compare the two alternatives according to the first property, and if they are equivalent, I use the second property, etc.). The relation \succsim_i is interpreted as the ranking in terms of salience of the different properties: at noon, the property ‘nice place to have lunch’ may typically be more salient than the property ‘touristic place’.

We also introduce i 's belief about j 's checklist, denoted \mathcal{C}_{ij} (i.e. the checklist that i believes j has in mind), and i 's belief about the *common* checklist \mathcal{C}_{iC} , the checklist that i believes she 'shares' with the other players, i.e. a checklist that i believes is common belief among the players. This common checklist is not necessarily the checklist of a given individual, it can be seen as the checklist of the 'group' of players, i.e. what distinguishes them as a particular group. While \mathcal{C}_i and \mathcal{C}_{ij} are necessary to define how i will represent the initial problem as a small world, \mathcal{C}_{iC} will determine the players' prior beliefs about the outcome of the game.

3.2 From large to small worlds

As discussed above, \mathcal{C}_i gives a procedure to partition the set of GW-acts. The player then derives her set of strategies from the set of GW-acts she is aware of. A *strategy* is a set of GW-acts for player i such that, for all the GW-acts of players in $-i$, the corresponding set of combinations of GW-acts has the same properties. In other words, the partition defined above allows the player to define indifference classes over GW-acts. For instance, if I am aware of the outcome 'we both meet at the Eiffel tower', I define 'go to the Eiffel tower' as a strategy, though it corresponds to a large set of GW-acts (i.e. all the GPS coordinates and altitude corresponding to the Eiffel tower). The question of how to choose a GW-act among this indifference class is left for after, once the small world has been solved – rather than defining immediately my precise choice, I treat the problems sequentially, by first finding a good spot in Paris, and then a good spot at the Eiffel tower. This second problem is the large world that remains to be solved once I am at the Eiffel tower (and again, I will frame it as a small world, and look for salient outcomes such as 'meeting at the south leg', 'meeting at the ticket office', etc.).

Once the player has framed her own set of strategies, she will imagine how the other players frame their own set of strategies. She will *simulate* the framing of player j by following the exact same procedure than the one described above (i indeed assumes by default that her own reasoning is a good predictor of the reasoning of j – this constitutes the basic principle of simulation thinking), while taking \mathcal{C}_{ij} as the input. If I am asked to meet a foreign tourist in Paris, there is little reason to expect that she will be aware of the outcome 'meet at Chatelet les Halles', but only of the outcome 'meet at the Eiffel tower'.

If the other individual is Parisian, it seems more reasonable to assume that she will be aware of both outcomes, though she may not be aware of the outcome ‘meet at Paris School of Economics’. By simulation, player i repeats the same operation than for her set of strategies, but by considering a partition of the GW-acts only with her belief about j ’s checklist \mathcal{C}_{ij} .^{xiv} We now have a set of strategy for player i and for the other player.

Given the awareness set of player i and the set of strategy profiles X defined above, we now investigate the payoff that player i attributes to each strategy profile. When considering all the possible strategy profiles (which, by construction, do not satisfy the same properties), the player will rank them according to whether they satisfy the properties she finds the more salient. This means that the ranking of the strategy profiles \triangleright induced by the checklist \mathcal{C}_i represent the relative salience of the different outcomes the player is aware of.

Proposition: Let X be the set of strategy profiles as defined above, and \mathcal{C} a checklist. There exists a strictly increasing payoff function $\Pi: X \mapsto \mathbb{R}$ such that $\Pi(x) > \Pi(x')$ if and only if $x \triangleright x'$.^{xv}

The proposition above means that we can define a payoff function from our basic ingredients, i.e. that given the player’s checklist and her beliefs about the checklist of the other, we can define (i) a strategy space, and (ii) a set of payoff functions for each player. This means that we have derived a small world representation from the initial large world.

An interesting point of the proposition is that the payoff function is strictly increasing: in other words, we cannot find two profiles x and x' such that their payoff is the same. This is due to the way the small world representation was build: if two GW-acts satisfy the exact same properties, then they are virtually undistinguishable from the perspective of player i . It is therefore possible to treat those two acts as part of the same ‘event’ (in Savage’s terminology), which implies that two identical acts should be considered identically by the decision maker. This directly echoes the principle of ‘Equal treatment of payoff-equivalent strategies’ of Mariotti (1997): if two strategies yield the exact same payoff for player i , then there is no reason for a player j to consider that player i may attach a higher probability for one of the strategies than the other. Two actions yielding the same payoff can therefore be treated as the *same* strategy.^{xvi}

3.3 Focal points

We can now define our notion of focal points. Intuitively, a focal point refers to a strategy profile that is more ‘salient’ than the others. Since the payoff function defined above determines the profile that satisfies the properties that the agent finds salient, the function $\Pi_i(x)$ gives us a measure of the salience of each profile, from the perspective of a player i .

Definition 1. Let $G_i(\mathcal{C}_i) = \langle N, X, \Pi \rangle$ be the small world defined by player i representing the strategic interaction GW , with X the set of strategy profiles, and $\Pi_j, \forall j \in N$, the payoff function of player j , when the set of strategies and payoff functions are defined with respect to i ’s checklist. The strategy profile $x^* \in X$ that maximises $\Pi_i(x)$ is called i ’s *individual focal point*.

Definition 2. Let $G_i(\mathcal{C}_{iC}) = \langle N, X, \Pi \rangle$ be the small world defined by player i representing the strategic interaction GW , with X the set of strategy profiles, and $\Pi_j, \forall j \in N$, the payoff function of player j , when the set of strategies and payoff functions are defined with respect to i ’s belief of the common checklist. The strategy profile $x^* \in X$ that maximises $\Pi_{iC}(x)$, the function that represents \mathcal{C}_{iC} is called i ’s *subjective social focal point*.

Definition 3. Let $G_i(\mathcal{C}_{iC})$ be the small world defined by player i representing the strategic interaction GW , with X^i the set of strategy profiles framed by player i . A strategy profile x^* is an *objective social focal point* if and only if:

- (i) $\forall i \in N, x^* \in X^i$
- (ii) $\forall i \in N, x^*$ is a subjective social focal point of the game $G_i(\mathcal{C}_{iC})$

The difference between the individual and subjective social focal points captures the idea of strategic framing, since the individual does not only pay attention to her own payoff (i.e. the outcomes that seem salient from *her* perspective), but restricts it to what she thinks is the shared perception among the group. Note also that, since the payoff functions are *strictly* increasing by construction of the small world itself, there necessarily exists a unique individual focal point, as well as a unique subjective social focal point. The objective social focal point exists if and only if the strategy profile in question is conceivable by all the players (for instance, a tourist will not be aware that Châtelet les

Halles is a possible profile), and that they all identify this profile as their subjective social focal point – which may still be different from their individual focal point.

The definition of a social focal point may look restrictive, since it could seem that we require it to be a payoff-dominant outcome (like HH in Hi-Lo or cooperation in a Stag Hunt). It must however be emphasised that the payoffs we traditionally associate in a Hi-Lo or Stag Hunt do not capture all the features of our notion of ‘payoff’ (which is multidimensional and captures the different properties of the checklist). This means that, when we ask players in a lab experiment to choose based on a payoff matrix, this is still a large world: we do not consider how the players will interpret what seem to be irrelevant features of the problem, such as the label of the strategies. We emphasise that the matrix representation we use in practice are not necessarily the representations used by the player – and that the labels we may arbitrarily add to the strategies for sakes of readability, cannot be used by the players we model (this is a label for *us*, the theorists, and not something useable by the player). A payoff dominant outcome in our framework is not a payoff that gives the highest monetary rewards to both players: it is the most salient profile, once we considered all the various aspects of the problem (individual monetary gains, the repartition of those gains, the label of the strategies, etc.). Our notion of ‘payoff’ is close to a behaviouristic interpretation in the sense that it represents ‘all-things-considered preferences’ and is thus tightly related to the individual’s choice, though it is used as an input of decision-making: we do not define the payoffs from the behaviour of the players, but from their mental states.^{xvii} There is then a close relation from our notion of payoff to the choice of the player – as the *only* motivating reason for the player to act – but it is not tautological as with behaviouristic preferences. The fact that there exists a payoff-dominant profile simply means that there is a form of alignment of the players’ interests, because they both identify the same outcome as the most salient.

Another possible concern is that our definition suggests that any game can have a social focal point: even though this assumption is realistic in matching games, it becomes much less convincing in e.g. zero-sum games. This is because – to keep the mathematical presentation in its simplest form – we did not consider for now the possibility of mixed strategies. We could echo here the distinction between the ‘massaged belief hierarchy’ and ‘subjective belief equilibrium’ in Larrouy and Lecouteux (2017): the massaged belief hierarchy corresponds to the equilibrium beliefs (possibly in correlated strategies) that

the players may hold, while the subjective belief equilibrium is only considered as the equilibrium in pure strategies when the players correctly anticipated the behaviours of others. A simple extension of the definitions above would be to allow correlated strategies for the definition of subjective social focal points (I can for instance believe in a matching game that we should alternate between two equilibria over time), while keeping a definition in pure strategies for objective social focal points (since they must actually occur, to be able to shape the beliefs of the other players). The implication would be that there could always exist a subjective social focal point for zero sum games (e.g. the mixed Nash equilibrium), while there is none objective social focal point.

3.4 Mindshaping and the formation of prior beliefs

We have seen thus far how players can frame a large world into a small world, thanks to their own checklist (this part of framing is largely unconscious, as in Bacharach's variable frame theory), and to their beliefs about the checklist of the other. We have also seen that we can always define an individual and a social subjective social focal point in this framework, by construction of the payoff functions Π_i and Π_{iC} . The question that must now be solved is how players choose, once they have framed the initial problem as a small world. A recurrent point of discussion is indeed whether it is rational or not to play one's part of a focal point. The attempt to rationalise focal points play on purely individualistic grounds usually faces an infinite regress problem (as it is not clear why it should be commonly believed that we all coordinate on the focal point rather than another profile; see e.g. Sugden 1991). We must usually add an *ad hoc* principle to justify the selection of Pareto-superior profiles, such as Luce and Raiffa (1957) 'solution in the strict sense', Harsanyi and Selten (1988) payoff-dominance principle, Bacharach (1993) 'principle of coordination', or Janssen (2006) 'principal of individual team member rationality'. An alternative way is to allow for a form of collective agency, and to consider that players may think of themselves as part of a group, who should play their part of the collectively rational profile – such theories of *team reasoning* have been proposed by Sugden (1991, 1993) and Bacharach (1999, 2006).^{xviii} Another road, which avoids possible debates about the nature of collective agency, is to consider the case of *evidential reasoning*, and that players may typically condition their beliefs about the others' actions on their own actions (but this remains a controversial position, see Weirich (2016) for a discussion).

The solution we suggest is that subjective social focal points offer a basis for the prior beliefs held by the players. This does not suggest that playing focal points is ‘rational’, but merely that, since we tend to spontaneously identify the focal point as the obvious solution of the game (by definition of the focal point, as something that spontaneously comes to your mind), it seems empirically relevant to consider that our prior gut feelings when we come to play a game are derived from this first intuition. Our argument is therefore that, if a profile – for any material, cultural, historical, or other reason – seems to be the one that ‘we’, the group of players involved in the game (Parisians, colleagues, everybody...), identify as the obvious one, then we should start any game-theoretic reasoning based on this prior belief. Given their prior beliefs, the players may then start to think strategically by simulating the reasoning of the other players, as in Larrouy and Lecouteux (2017). They should check that their prior belief is consistent in a Bayesian sense (I cannot believe that you play for instance a dominated strategy, nor that you could believe that I play a dominated strategy). This process of simulation will lead the players to ‘massage’ their prior beliefs until they become consistent – and the existence of a massaged belief hierarchy derived from this prior belief is ensured by the finiteness of the small world. Once their beliefs are consistent, the players simply maximise their expected payoff (defined thanks to their checklist \mathcal{C}_i).^{xix}

As soon as we recognise that players are socially-embedded agents, who necessarily (voluntarily or not) belong to certain groups, directed by various norms and rules (translated in our framework as various *common checklists*), then considering that those various social influences shape our prior beliefs seems to be a reasonable assumption. In this perspective, we could relatively easily model the role of mindshaping in a game-theoretic setting. Since the prior belief of player i corresponds to the social subjective focal point (the profile x that maximises the function Π_{iC}), it will be sufficient to study the evolutionary dynamics of prior beliefs, and to consider that changes in prior beliefs are the result of an underlying change in \mathcal{C}_{iC} , and therefore of the rules and norms that i believes are shared by the group of players. An objective social focal point would be reached only once the prior beliefs of the players become aligned, and all identify the same strategy profile as their subjective social focal point.^{xx}

4. Normative implications

The aim of this paper was to offer a theory of coordination that considers the role of the context within which the individuals interact. This required attacking the problem of choosing in large worlds, i.e. of explicitly dealing with open rather than closed systems. This allowed us to highlight the role of the processes of mindshaping in the formation of the preferences and beliefs of the individuals, both as influencing the way the players conceptualise their decision problems (when framing the initial large world as a self-contained small world), and as influencing their gut feelings, and hence their prior beliefs in the small world representation of the decision problem. Unlike within the standard approach to coordination games, the players can identify *ex ante* a simple solution to the game, and they need not form complex belief hierarchies prior to their choice. A possible methodological argument in favour of this departure from standard Bayesian game theory is to consider that game theoretic reasoning is not conscious deliberation (Gold 2018, pp.340-342), and that beliefs are formed as a way to rationalise *a posteriori* our actions. It is because players have been exposed to the same social dynamics involved in the processes of mindshaping that they are able to coordinate on focal points, and beliefs merely play a role of *ex post* justification. This suggests that, unlike within Bayesian game theory, beliefs could be treated as an output, rather than input, of successful coordination (Guala, forthcoming).

Our study has been largely descriptive until now, though the way we integrated the notion of ‘context’ in the analysis raises important normative issues, in particular with respect to the recent literature on behavioural welfare economics. Traditional welfare economics assumes that individuals have stable and context-independent preferences, and uses preference satisfaction as a normative criterion. Behavioural economics have called this assumption into question, which raises fundamental problems for normative economics: if people’s preferences are likely to change over time, or to depend on apparently irrelevant aspects of the choice situation, can we still form normative judgments about people’s choices based on their revealed preferences? The problem of how to reconcile normative and behavioural economics (labelled by McQuillin and Sugden (2012) as ‘the reconciliation problem’) is that economists need to develop an alternative normative criterion to the standard preference satisfaction paradigm. The main approach developed thus far, *behavioural welfare economics*, consists in treating

departures from conventional rational choice theory as mistakes, and uses the satisfaction of the ‘true’ preferences of the individuals – the preferences they would have revealed, were they able to reason correctly – as a normative criterion.

This view of the individual as an ‘inner rational agent trapped in an outer psychological shell’ however presents various methodological limitations and lacks serious psychological foundations (Infante *et al* 2016). More specifically, psychology is treated as a set of forces that interferes with the latent rationality of the individual, who is refrained from behaving as the rational agent she intends to be. This type of discourse tends to ‘pathologize’ the individual, whose defective psychology is identified as the cause of all her troubles. Obesity or poverty are indeed due to a lack of self-control, as psychological biases prevent people from resisting the temptation of sugar binges, or the desire to spend frenetically their money. While this discourse tends to justify a new form of ‘behavioural’ paternalism, by extending ‘the paternalistically protected category of “idiots” to include most people, at predictable times’ (Camerer *et al* 2003, p.1218), it also remains silent about the social dynamics that characterise the environment within which the individuals evolve – see Mehta (2013) on this point, according to whom ‘the upsurge of interest in these behavioural remedies distracts attention from facets of the individual’s environment, such as [...] the set of norms and beliefs underpinning the economic system’ (p.1244).

While behavioural welfare economics treats the ‘context’ as part of the biases that interfere with our latent rationality, we suggested to give it a much more substantive content, as an essential component of the processes of preference and belief formation. If we consider that preferences and beliefs are the product of the environment and the norms that guide the social interactions of the individuals – and therefore treat individuals as socially-embedded agents rather than isolated ‘faulty Econs’ – then we must explicitly discuss the processes of preference formation, and not limit ourselves with a criterion of preference satisfaction (Lecouteux 2015). Behavioural welfare economics seems normatively warranted if humans are defective *homo oeconomicus*. Behavioural evidence however seem to suggest not that people systematically make mistakes (and fail to satisfy some inner, true, preferences) but rather that the environment is likely to influence their preferences, typically through the processes of mindshaping. Rather than individualising the causes of the agents’ behaviours, we should recognise the influence of social dynamics on the formation of individual preferences and beliefs, and thus of the

normativity of the rules and institutions that guide our interactions – we can indeed wonder whether it is normatively desirable to help people to satisfy their preferences, if those preferences were shaped by influences that the individual would reject, if she were aware of them (such as gender stereotypes on the choices of career or study). Our analysis suggests that social norms and focal points, in addition to facilitate coordination, may also greatly influence how we frame our decision problems. This means that the locus of normative analysis should not necessarily be how to steer people into the ‘right’ direction to satisfy what we think they truly want, but rather how to ensure the autonomous formation of individual preferences and beliefs, by considering the influence of social dynamics and norms on the individuals.

References

- Anscombe, F. J., & Aumann, R. J. (1963). A definition of subjective probability. *Annals of mathematical statistics*, 34(1), 199-205.
- Aumann, R. J., & Dreze, J. H. (2009). Assessing strategic risk. *American Economic Journal: Microeconomics*, 1(1), 1-16.
- Bacharach, M. (1993). Variable Universe Games. In *Frontiers of game theory*.
- Bacharach, M. (1999). Interactive team reasoning: A contribution to the theory of co-operation. *Research in economics*, 53(2), 117-147.
- Bacharach, M. (2006). *Beyond individual choice: teams and frames in game theory*. Princeton University Press.
- Bacharach, M., & Bernasconi, M. (1997). The variable frame theory of focal points: An experimental study. *Games and Economic Behavior*, 19(1), 1-45.

- Binmore, K. (1988). Modeling rational players: Part II. *Economics & Philosophy*, 4(1), 9-55.
- Binmore, K. G. (1993). De-Bayesian Game Theory. In Binmore, K. G., Kirman, A. P., & Tani, P. (Eds.). *Frontiers of game theory*. MIT Press.
- Binmore, K. G. (2009). *Rational decisions*. Princeton, NJ: Princeton University Press.
- Camerer, C., Issacharoff, S., Loewenstein, G., O'donoghue, T., & Rabin, M. (2003). Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism". *University of Pennsylvania law review*, 151(3), 1211-1254.
- Casajus, A. (2000). Focal points in framed strategic forms. *Games and Economic Behavior*, 32(2), 263-291.
- De Jongh, D., & Liu, F. (2009). Preference, priorities and belief. In *Preference Change* (pp. 85-107). Springer, Dordrecht.
- Dietrich, F., & List, C. (2013). Where do preferences come from?. *International Journal of Game Theory*, 42(3), 613-637.
- Dietrich, F., & List, C. (2016). Mentalism versus behaviourism in economics: a philosophy-of-science perspective. *Economics & Philosophy*, 32(2), 249-281.
- Giocoli, N. (2003). *Modeling rational agents: From interwar economics to early modern game theory*. Edward Elgar Publishing.
- Gold, N. (2018). Team Reasoning and Spontaneous Collective Intentions. *Revue d'économie politique*, 128(3), 333-353.
- Gold, N., & Colman, A. M. (2018). Team Reasoning and the Rational Choice of Payoff-Dominant Outcomes in Games. *Topoi*, 1-12.

- Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford, NY: Oxford University Press.
- Goyal, S., & Janssen, M. (1996). Can we rationally learn to coordinate?. *Theory and Decision*, 40(1), 29-49.
- Guala, F. (2018). Coordination, Team Reasoning, and Solution Thinking. *Revue d'économie politique*, 128(3), 355-372.
- Guala, F. (forthcoming), Solving the Hi-lo Paradox: Equilibria, Beliefs, and Coordination, in *Minimal Cooperation and Shared Agency*, edited by A. Fiebich. Dordrecht: Springer.
- Harsanyi, J. C., & Selten, R. (1988). A general theory of equilibrium selection in games. *MIT Press Books*.
- Hédoin, C. (2014). A framework for community-based salience: common knowledge, common understanding and community membership. *Economics & Philosophy*, 30(3), 365-395.
- Infante, G., Lecouteux, G., & Sugden, R. (2016). Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology*, 23(1), 1-25.
- Janssen, M. C. (2001). Rationalizing focal points. *Theory and Decision*, 50(2), 119-148.
- Janssen, M. C. (2006). On the strategic use of focal points in bargaining situations. *Journal of Economic Psychology*, 27(5), 622-634.
- Kahneman, D. and A. Tversky, eds. (2000). *Choices, values, and frames*.

- Larrouy, L. (2018). The ontology of Schelling's "Theory of Interdependent Decisions". *The Individual and the Other in Economic Thought: An Introduction*.
- Larrouy, L., & Lecouteux, G. (2017). Mindreading and endogenous beliefs in games. *Journal of Economic Methodology*, 24(3), 318-343.
- Lecouteux, G. (2015). In search of lost nudges. *Review of Philosophy and Psychology*, 6(3), 397-408.
- Lecouteux, G. (2018a). Bayesian game theorists and non-Bayesian players. *European Journal of the History of Economic Thought*. In press.
- Lecouteux, G. (2018b). What does "we" want? Team Reasoning, Game Theory, and Unselfish Behaviours. *Revue d'économie politique*, 128(3), 311-332.
- Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. Courier Corporation.
- Mandler, M., Manzini, P., & Mariotti, M. (2012). A million answers to twenty questions: Choosing by checklist. *Journal of Economic Theory*, 147(1), 71-92.
- Mariotti, M. (1995). Is Bayesian rationality compatible with strategic rationality?. *The economic journal*, 1099-1109.
- Mariotti, M. (1997). Decisions in games: why there should be a special exemption from Bayesian rationality. *Journal of Economic Methodology*, 4(1), 43-60.
- McQuillin, B., & Sugden, R. (2012). Reconciling normative and behavioural economics: the problems to be solved. *Social Choice and Welfare*, 38(4), 553-567.
- Mehta, J. (2013). The discourse of bounded rationality in academic and policy arenas: pathologising the errant consumer. *Cambridge journal of economics*, 37(6), 1243-1261.

- Mehta, J., Starmer, C., & Sugden, R. (1994a). The nature of salience: An experimental investigation of pure coordination games. *The American Economic Review*, 84(3), 658-673.
- Mehta, J., Starmer, C., & Sugden, R. (1994b). Focal points in pure coordination games: An experimental investigation. *Theory and Decision*, 36(2), 163-185.
- Orléan, A. (2004). What is a collective belief?. In *Cognitive economics* (pp. 199-212). Springer, Berlin, Heidelberg.
- Ramsey, F. P. (1926). The foundations of mathematics. *Proceedings of the London Mathematical Society*, 2(1), 338-384.
- Ross, D. (2007). H. sapiens as ecologically special: what does language contribute?. *Language sciences*, 29(5), 710-731.
- Savage, L. J. (1954). *The foundations of statistics*. Courier Corporation.
- Schelling Thomas, C. (1980) [1960]. The strategy of conflict. *Harvard University Press*.
- Sugden, R. (1991). Rational choice: a survey of contributions from economics and philosophy. *The economic journal*, 101(407), 751-785.
- Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behavior. *Social philosophy and policy*, 10(1), 69-89.
- Sugden, R. (1995). A theory of focal points. *The Economic Journal*, 533-550.
- Sugden, R. (2015). Team reasoning and intentional cooperation for mutual benefit. *Journal of Social Ontology*, 1(1), 143-166.

von Neumann, J. (1983 [1931]), The formalist foundations of mathematics, in Benacerraf, P. and Putnam, H. (eds), *Philosophy of Mathematics. Selected Readings*, Cambridge: Cambridge University Press, 61-65.

Von Neumann, J., & Morgenstern, O. (1953). *Theory of games and economic behavior*. 3rd edition, Princeton university press.

Weibull, J. W. (1997). *Evolutionary game theory*. MIT press.

Weirich P (2016) Causal Decision Theory. In: Zalta NE (ed) The Stanford encyclopedia of philosophy (Winter 2016 Edition).

<https://plato.stanford.edu/archives/win2016/entries/decision-causal/>

Zawidzki, T. W. (2013). *Mindshaping: A new framework for understanding human social cognition*. MIT Press.

Zawidzki, T. W. (2018). Mindshaping. In *The Oxford Handbook of 4E Cognition*.

ⁱ The two main approaches are the ‘theory-theory’ and the ‘simulation theory’. See Larrouy and Lecouteux (2017) and Guala (2018) for a brief presentation and a discussion in a game theoretical context.

ⁱⁱ In addition to the points discussed here, see also Mariotti (1995, 1997) on the problems of extending Savage’s axioms to game theory.

ⁱⁱⁱ A conceptually related framework would be Dietrich and List (2013) model of ‘motivationally salient properties’, with a restriction to lexicographic weighing relations over property combinations, as in de Jongh and Liu (2009).

^{iv} Note that, in the process described above, it is very likely that the small world representation of a player may fail to take into account some important features of the problem. In the case of Brexit, we can mention for instance the property of ‘deterrence’ – that the UK fails to recognise as salient for the EU – or the ‘status of Northern Ireland’, which has been set aside in our small world representation. The players may also be unaware of some strategies (e.g.

finding a new type of arrangement different from merely remaining or leaving the single market).

- ^v An important implication of defining payoffs this way is that they can be observed *prior to the choice*. This means that we can experimentally test the theory that players behave as if they intend to maximise their expected payoff, while it is tautologically satisfied – by definition of the utility function itself – in the behavioural approach (see Lecouteux 2018b).
- ^{vi} For arguments in favour of mentalism in economics, see Dietrich and List (2016) and Lecouteux (2018b), who both argue that it is the basis of better scientific practices.
- ^{vii} We do not mean here that there is no ontological difference at all between small and large worlds, but that even some closed systems are large worlds – and any open system may be presumably considered as a large world.
- ^{viii} For other aspects of Schelling’s account of focal points, see Larrouy (2018).
- ^{ix} This distinction between individual and social focal points means that an individual can differentiate her own frames, intentions or beliefs, from those she believes are common among the group (Orléan 2004). Framing becomes strategic, since the player can distance herself from her own initial framing, and strategically reframe the game.
- ^x This is known as the ‘egocentric bias’ (see Goldman, 2006, pp 177-179), and the observation of such bias can be used as evidence of simulation thinking.
- ^{xi} Similarly, Ross (2007) suggests that it is only because some social devices shape people’s mind so that they become more alike that coordination can occur.
- ^{xii} As emphasised by Casajus (2000, 264-265), this two-step process of describing how players (i) frame the game, and (ii) choose once the problem has been framed, ‘is an important step towards the rationalisation of salience’, though ‘this kind of rationalisation of choosing the salient is to be distinguished from the formalisation of salience itself’ (see also Goyal and Janssen 1996).
- ^{xiii} We do not need that properties map into the set of integers, but more generally into a set that can be well-ordered by the property
- ^{xiv} In principle, player *i* can also consider that there are properties that *j* is aware of, but not her – e.g. ‘meet at Institut Marie Curie’, which would be salient for some doctors but probably not for an economist. We will not discuss such cases here, and always assume that *i* reasons as if she necessarily knew at least as much as the other players.
- ^{xv} The proof of the proposition follows directly from the theorem 2 of Mandler *et al* (2012) on ‘multivalued checklists’, since \triangleright is generated by the exact same procedure. The existence of the function is guaranteed by the countable number of properties, while the strict increasing is due to the construction of the small world, and the fact that two profiles

generating the same payoff are considered as the same strategy, by definition of the set of strategy profiles (which are indifference classes over GW-acts).

^{xvi} This is also consistent with the sequential approach of dealing with large worlds: once the small world has been solved (and a strategy x_i chosen), the player will realise that she must now refine her choice (e.g. go to the south leg of the Eiffel Tower), because the initial act included too many GW-acts.

^{xvii} A difficulty that may arise is then to properly calibrate the payoff function in the model, depending on the various properties that may motivate the players. In many lab experiments, we can probably confidently assume that the main property considered by the individuals is their monetary gains, but this assumption should not be systematic, in particular in games where concerns of fairness may be decisive.

^{xviii} See Lecouteux (2018b) for a review, and Gold and Colman (2018) for a discussion of team reasoning and the choice of payoff-dominant profiles.

^{xix} According to Larrouy and Lecouteux (2017), simulation thinking may be coherent with the rationalisation of non-ratifiable choices, meaning that this process may lead to the formation of action-dependent beliefs, which *in fine* could rationalise the choice of e.g. cooperation in a prisoner's dilemma, as soon as the underlying belief is that the players are both conditional co-operators.

^{xx} Formally, we can easily model the evolutionary dynamics as a multipopulation game, with each population corresponding to one individual with her collection of possible prior beliefs. Since stability requires the convergence of the prior beliefs to a single distribution, the underlying evolutionary game has a structure of a matching game, with several strict Nash equilibria. The only evolutionary stable states in this setting are the strict Nash equilibria (e.g. Weibull 1997), i.e. when all the players have the same prior beliefs. Given the simple structure of the game, we can expect a rapid convergence towards a stable state, depending on the initial conditions and the model considered for the replicator dynamics. Note also that the focal point in question should be a subjective belief equilibrium (Larrouy and Lecouteux 2017).

DOCUMENTS DE TRAVAIL GREDEG PARUS EN 2018
GREDEG Working Papers Released in 2018

- 2018-01** LIONEL NESTA, ELENA VERDOLINI & FRANCESCO VONA
Threshold Policy Effects and Directed Technical Change in Energy Innovation
- 2018-02** MICHELA CHESSA & PATRICK LOISEAU
Incentivizing Efficiency in Local Public Good Games and Applications to the Quantification of Personal Data in Networks
- 2018-03** JEAN-LUC GAFFARD
Monnaie, crédit et inflation : l'analyse de Le Bourva revisitée
- 2018-04** NICOLAS BRISSET & RAPHAËL FÈVRE
François Perroux, entre mystique et politique
- 2018-05** DUC THI LUU, MAURO NAPOLETANO, PAOLO BARUCCA & STEFANO BATTISTON
Collateral Unchained: Rehypothecation Networks, Concentration and Systemic Effects
- 2018-06** JEAN-PIERRE ALLÉGRET, MOHAMED TAHAR BENKHODJA & TOVONONY RAZAFINDRABE
Monetary Policy, Oil Stabilization Fund and the Dutch Disease
- 2018-07** PIERRE-ANDRÉ BUIGUES & FRÉDÉRIC MARTY
Politiques publiques et aides d'Etat aux entreprises : typologie des stratégies des Etats Membres de l'Union Européenne
- 2018-08** JEAN-LUC GAFFARD
Le débat de politique monétaire revisité
- 2018-09** BENJAMIN MONTMARTIN, MARCOS HERRERA & NADINE MASSARD
The Impact of the French Policy Mix on Business R&D: How Geography Matters
- 2018-10** ADRIAN PENALVER, NOBUYUKI HANAKI, EIZO AKIYAMA, YUKIHIKO FUNAKI & RYUICHIRO ISHIKAWA
A Quantitative Easing Experiment
- 2018-11** LIONEL NESTA & STEFANO SCHIAVO
International Competition and Rent Sharing in French Manufacturing
- 2018-12** MELCHISEDEK JOSLEM NGAMBOU DJATCHE
Re-Exploring the Nexus between Monetary Policy and Banks' Risk-Taking
- 2018-13** DONGSHUANG HOU, AYMERIC LARDON, PANFEI SUN & THEO DRIESSEN
Compromise for the Per Capita Complaint: An Optimization Characterization of Two Equalitarian Values
- 2018-14** GÉRARD MONDELLO & EVENS SALIES
The Unilateral Accident Model under a Constrained Cournot-Nash Duopoly
- 2018-15** STÉPHANE GONZALEZ & AYMERIC LARDON
Axiomatic Foundations of a Unifying Concept of the Core of Games in Effectiveness Form
- 2018-16** CLAIRE BALDIN & LUDOVIC RAGNI
François Perroux : Echange pur contre échange composite - Controverses et enjeux de justice
- 2018-17** GUILHEM LECOUTEUX
What Does 'We' Want? Team Reasoning, Game Theory, and Unselfish Behaviours
- 2018-18** LUDOVIC DIBIAGGIO, BENJAMIN MONTMARTIN & LIONEL NESTA
Regional Alignment and Productivity Growth

- 2018-19** NICOLAS BRISSET
L'organe comme marchandise fictive : une mise en perspective
- 2018-20** DONGSHUANG HOU, AYMERIC LARDON, PANFEI SUN & HAO SUN
Procedural and Optimization Implementation of the Weighted ENSC Value
- 2018-21** FRÉDÉRIC MARTY
Intelligence Artificielle et Organisation Industrielle : quels enjeux pour l'économie numérique
- 2018-22** MICHELE PEZZONI, REINHILDE VEUGELERS & FABIANA VISENTIN
Is This Novel Technology Going to be a Hit? Antecedents Predicting Technological Novelty Diffusion
- 2018-23** JACQUES MAIRESSE & MICHELE PEZZONI
Novelty in Science: The Impact of French Physicists' Novel Articles
- 2018-24** GIUSEPPE ATTANASI, LAURA CONCINA, CAROLINE KAMATE & VALENTINA ROTONDI
Firm's Protection against Disasters: Are Investment and Insurance Substitutes or Complements?
- 2018-25** NICOLAS BRISSET
Bourdieu, neo-smithien ? À propos d'Anthropologie économique
- 2018-26** AGNÈS FESTRÉ & ALAIN RAYBAUT
Colonial Banks, Credit, and Circulation: The Example of Martinique 1848-1871
- 2018-27** FRÉDÉRIC MARTY
The Unspeakable One. De l'activation de la théorie des facilités essentielles dans l'économie numérique
- 2018-28** AGNÈS FESTRÉ
Hayek on Expectations: The Interplay between Two Complex Systems
- 2018-29** LAUREN LARROUY & GUILHEM LECOUTEUX
Choosing in a Large World: The Role of Focal Points as a Mindshaping Device