Silicone prophète

Prédictions sociales, anticipations, équilibres.

Henri Galinon
UCA - PHIER

Journée AFIA-SPS : IA et prédiction en sciences sociales
6 février 2020

Les deux défis de la prédiction en sciences sociales

- 1. Fiabilité
- 2. Responsabilité

Qu'est-ce que l'IA ? IA = puissance + délégation(s)

-> La puissance contribue à résoudre le problème de la fiabilité
-> La délégation pose deux problèmes :
    intelligibilité
    risque stratégique

Vue d’ensemble

1. Prédire/modéliser est difficile car les phénomènes sociaux sont complexes
2. Apport et limites de l’IA dans la prédiction des phénomènes sociaux

3. Spécificité des sciences sociales (par rapport à d’autres systèmes complexes) : la réflexivité.

-> quel impact sur la prédictibilité des phénomènes sociaux ?

-> quelle question éthique/politique spécifique pour la prédiction en sciences sociales ?

-> quel enjeu spécifique pour l’usage et la maîtrise de l’IA prédictive en sciences sociales ?

1.1 Viser la prédiction en sciences sociales ?

- Prédire :
  - prévoir l'issue d'un jet de dé vs. prévoir la fréquence des différentes issues.
  - Modèles déterministes ou stochastiques
- La prédiction en sciences sociales aujourd'hui :
  1. Peu de prédiction dans les sciences sociales vs. besoin de prédiction dans les applications
  2. Prédire est difficile : les systèmes sociaux sont complexes

--> Mais variété des tâches:

- Prédire la croissance au T1 2020 - ok agrégat avec observations riches et inertie importante
- vs. Prédire ex ante la date et l'ampleur de la prochaine crise économique – pas ok : événement rare par nature imprévisible
- Cas intéressants : quelles théories/modèles pour prédire la structuration d'un réseau social, prédire le succès d'un film, une cascade de tweets, etc.

1.2 Viser la prédiction: les apports de l'IA

- Avant : comprendre/expliquer (« insights ») mais peu de prédiction et de généralisations possibles, validation relâchée

  Au mieux expériences ramdomisées sur petits effectifs et recherche « significativité statistique » (psycho, économie)

- Les progrès par l'« IA » (« computational social science » Adamic et al., Conte et al., Salganik):

  « big data » + puissance de calcul = capacité de modélisation/validation démultipliée

  -> Possibilité d'utiliser des algo d'apprentissage d'hypothèses complexes (ML) et de valider sur des tâches de prédiction (type « cross validation »), ou de « fitter » sur les données des modèles complexes existants (percolation, attachement préférentiel, etc.), ou de faire de la simulation multiagents etc.

  → Les data et l'IA améliorent la capacité théorique de prédiction par rapport aux méthodes antérieurement utilisées

  → Leviers futurs : exploiter la possibilité d'expériences massives online, produire des données massives « propres » (« super collider ») Voir Salganik 2017, Watts 2016

1.3. Science et prédiction : Netflix Prize était-il un prix de sciences sociales ?

- Appel ouvert pour améliorer son algo de recommandation (2006)
- Défi : Prédire les évaluations de films par des utilisateurs
  - Data fournies: 100 millions d'évaluations de films (500 000 clients, 20 000 films)
  - Tâche prédictive : prédire les 3 millions d'évaluation existantes mais non fournies.
  - Incitation: 1 million $ si améliore de 10% l'algo maison (« cinematch »)

-> prédictif

-> validation par compétition sur une tâche (benchmarking)

1.3 « Boîte noire » et intelligibilité

- Si l'oracle de Delphes avait été plus transparent, Œdipe aurait sans doute pu éviter la tragédie.

Si on ne parvient pas à interpréter les modèles ML de prédictions alors :

-> difficile de pondérer, discuter, justifier à partir prédictions de « boîtes noires »

-> pas d'intelligibilité, pas de connaissance, pas de sciences sociales ?

Conflit explication/prédiction ?

Le message des Computational Social Sciences :

On veut les deux : prédiction + intelligence

2.1 De la prédiction difficile à la prédiction impossible

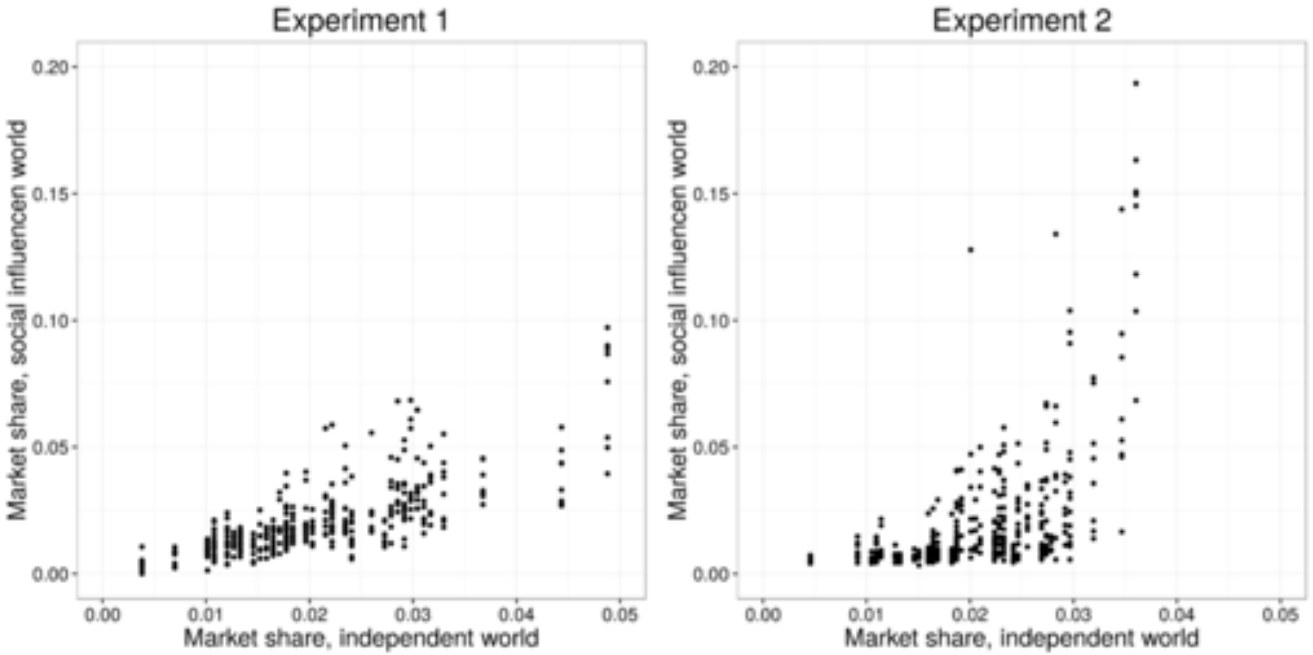
Le Music Lab de Salganik, Dodds, Watts 2006

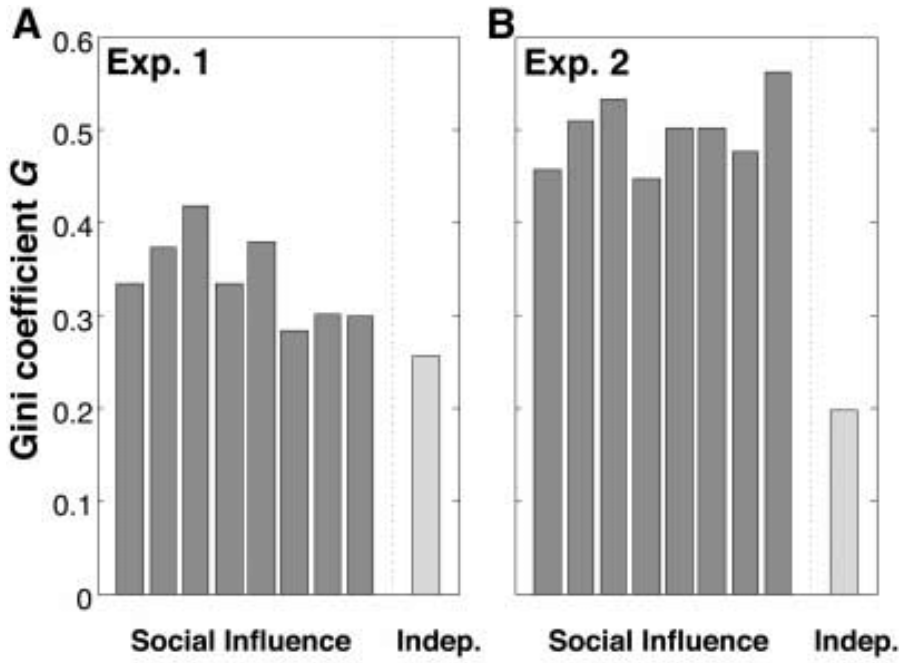
- 14000 participants, 48 morceaux de musique à évaluer et télécharger
- 2 conditions :

Condition Indépendance : la liste des morceaux présentée sans infos agrégées sur les éval. ni le nbre de téléchargement

Condition Influence : Exp 1 : signal note moyenne et nbre de téléchargements / Exp 2 : Exp.1 + présentation des morceaux en liste ordonnée par popularité

-> dans la condition Influence les participants sont répartis en huit « mondes » qui fonctionnent indépendamment (mêmes conditions initiales)





2.2 Estimer les limites de la prédictibilité ?

- Martin & al. 2016

Peut-on prévoir le succès d'un tweet ?

(= le nombre de retweets)

Idée

Succès = f(qualité intrinsèque) + chance (bruit)

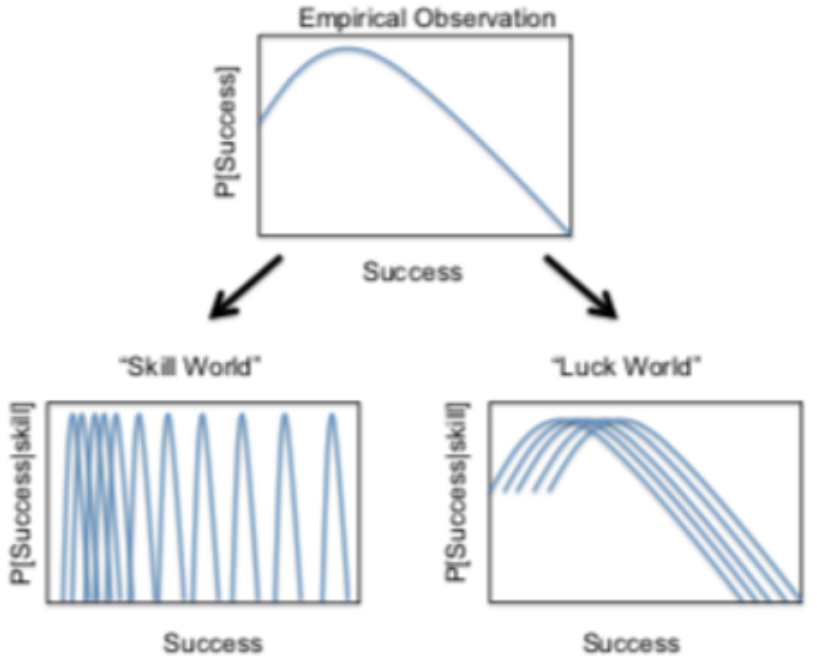


Figure 1: Schematic model illustrating two stylized explanations for an empirically observed distribution of success.

2.2 Estimer les limites de la prédictibilité ?

- Martin & al. 2016. Etude empirique + simulations

Thèse :
Variabilité intrinsèque aux cascades tweeter et existence d'une limite théorique à la prédictibilité du succès des tweets

Model	*Tweet time*	*Domain*	*Spam score*	*Category*	*Tweet topic*	*Past url success*	*User time*	*Followers*	*Friends*	*Statuses*	*User topic*	*Past user success*	*Topic interaction*
1. Basic content	✓	✓	✓	✓									
2. Content, topic	✓	✓	✓	✓	✓								
3. Content, past succ.	✓	✓	✓	✓	✓	✓							
4. Basic user							✓	✓	✓	✓			
5. User, topic							✓	✓	✓	✓	✓		
6. User, past succ.							✓	✓	✓	✓	✓	✓	
7. Content, user	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
8. All	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 2: Features used in different models for cascade size prediction.

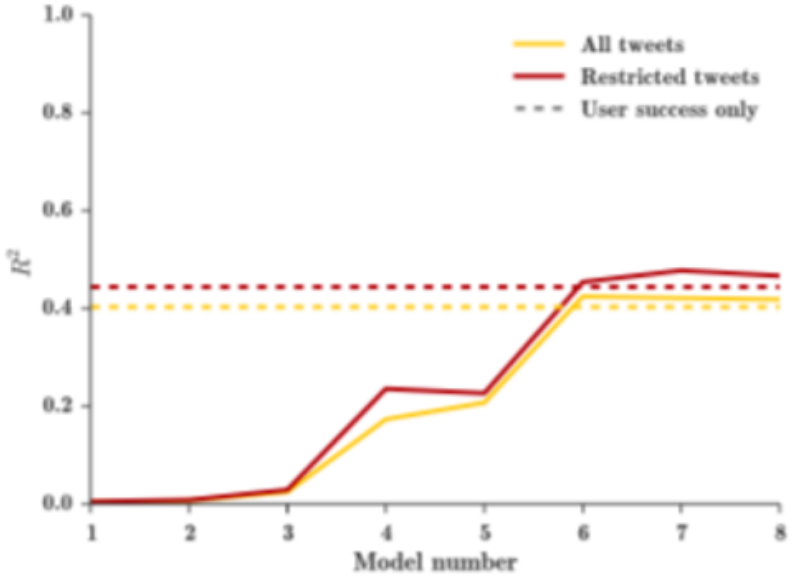


Figure 4: Prediction results for models using different subsets of features. R^2 increases as we add more features, but only up to a limit. Even a model with all features explains less than half of the variance in cascade sizes.

2.3 (Autres facteurs d'imprédictibilité)

- Sensibilité aux conditions initiales (Poincaré, Duhem, th. du chaos)
- Structure de disponibilité de l'info sur le système à prédire (marche au hasard des prix sur les marchés financiers efficients)
- ...
- Conflit stratégique du prédicteur et du prédit (Scriven, Lewis, Laegerspetz)

Prédicteur \ Joueur	**Faire A**	**Faire B**
Prédire A	(2, 0)	(0,2)
Prédire B	(0,2)	(2,0)

Si le joueur a une puissance de calcul supérieure à celle du prédicteur,
il est imprévisible pour le prédicteur

3. La réflexivité

- Prédire des comportements pour anticiper, anticiper pour adapter son comportement, comportements adaptatifs individuels qui affectent la réalisation de la prédiction sociale (sans « intention collective » de l'affecter)

Note : pouvoir causal non pas de la prédiction elle-même mais de sa diffusion (on ne parle pas de martiens qui prédisent phénomènes sociaux terriens sans contact avec eux)

- Exemples

SFP – prophétie auto-réalisatrice	**SDP – Prophétie auto-réfutante**
1. p a été prédit	1. p a été prédit
2. p	2. non-p
3. Si p n'avait pas été prédit, alors non-p serait le cas	3. Si p n'avait pas été prédit, alors p serait le cas

- Spéculation financière (bulles), confiance, mœurs (2 bises ou 3 bises ?), etc.

→ La plupart des lois de la vie sociale telles qu'elles sont ordinairement formulées semblent potentiellement vulnérables des effets de réflexivité (mutabilité)

3.1 Réflexivité et sciences sociales

La réflexivité distingue les sciences sociales des sciences de la nature.

Central pour comprendre ce que c'est que de faire des sciences sociales ?

- Merton (36, 48) : « Self-fulfilling prophesy » (SFP)
- Giddens (79) : sociology as « critical theory »
- Barnes, Callon , MacKenzie (80'-00') : Performativité, Bootstrapped induction
- Economie des conventions, économie hors équilibre « fondamentaux », réflexivité de l'économie (Soros) ...
- Dupuy : « auto-transcendance »

- **Merton** :

« There is one other circumstance, peculiar to human conduct, which stands in the way of successful social prediction and planning. Public prediction of future social developments are frequently not sustained precisely because the prediction has become a new element in the concrete situation thus tending to change the initial course of development. This is not true of prediction in fields which do not pertain to human conduct. […]. Marx's prediction of […] increasing misery of the masses [resulted in the spread of organization of labor thus] slowing up […] the developments which Marx had predicted.» (Merton, « The unanticipated consequence of purposive social action » p. 904)

« The parable tells us that public definition of a situation (propheties or prediction) become an integral part of the situation and thus affect subsequent developments. This is peculiar to human affairs. It is not found in the world of nature. Predictions of the return of Halley's comet do not influence its orbit. But the rumored insolvency of Millingville's bank did affect the actual outcome. The prophecy of collapse led to its own fulfillment » (Merton, « The self fulfilling prophecy », p.195)

- Giddens 1979, *Central Problems in Social Theory* :

« Rather than attempting to marginalise, and treat purely as a 'problem', the potential incorporation of social scientific theories and observations within the reflexive rationalisation of those who are their 'object' - human agents - we have to treat the phenomenon as one of essential interest and concern to the social sciences. For it becomes clear that every generalisation or form of study that is concerned with an existmg society constitutes *a* ***potential intervention within that society****:* and this leads through to the tasks and aims of sociology ***as critical theory*** *»* (p. 244-245, Giddens souligne)

- Henshel 1982:

TABLE I *Major areas with self-fulfilling prophecies research* (after Henshel (1978), prepared jointly by Henshel and R. K. Merton)*

I. *Race and ethnic relations*
- (1) Minority stereotypes as self-fulfilling
- (2) Ecological invasion-succession ('block-busting')

II. *Deviant behaviour and social control*
- (1) Labelling of deviants perpetuating deviancy
- (2) Paranoid delusion as self-fulfilling

III. *Models of 'human nature' as self-fulfilling*

IV. *Education*
- (1) Self-fulfilling aspects of teacher expectancy
- (2) Self-fulfilling aspects of school testing, tracking, streaming
- (3) Self-image and performance as an SFP

V. *Scientific inquiry*
- (1) Investigator expectancy
- (2) Subject expectancy
- (3) Placebo effect

VI. *Politics, law, and international relations*
- (1) Predictions of voting behaviour
- (2) Escalation and conflict resolution
- (3) Administration of the law

VII. *Economics*
- (1) Market fluctuation
- (2) Inflation and depression spirals
- (3) Trend projection as self-fulfilling
- (4) Occupational stereotypes

VIII. *Religion*
- (1) Millenarianism, mysticism
- (2) Faith healing as SFP

*For references for each category see R. L. Henshel, 'Self-Altering Predictions', in Fowles, *Handbook of Futures Research*, Westport, Conn., Greenwood Press, 1978, pp. 99-123.

3.2 Réalité expérimentale des SFP

- En laboratoire : Salganik & Watts 2007 Music Lab exp. :

-> la manipulation des classements de popularité (affichage du nombre de téléchargements) manipule le niveau de popularité (nombre de téléchargements)

(-> l'effet s'estompe avec le temps)

- Hors labo : étude de MacKenzie 2006 sur les modèles de Black-Sholes-Merton de pricing des dérivés sur les marchés d'options.

- Voir Biggs 2011 pour d'autres références

3.3 Périmètre d’exclusion de la réflexivité

Les limites de la réflexivité (inspiré de Hershel 1982 et Lagerspetz 1988) et invariants sociaux :

- Le comportement prédit est contraint extérieurement, non rationnel, compulsif par nature
- Les prédictions ne sont connues que de personnes qui n’ont aucun interaction directe ou indirecte avec les groupes sociaux concernés par la prédiction (= martiens)
- Les prédictions sont jugées non crédibles ou sans sans intérêt
- Les sujets ont une stratégie dominante dans la situation prédite
- Les prédictions sont court-termistes (horizon temporel qui précède le moment où leur validité sera affectée par les comportements adaptatifs déclenchés directement ou indirectement par la prévision)

3.4 Réflexivité + fiabilité

- Le problème (stylisé) de Bison Futé. Hypothèse : la décision de prendre la voiture dépend de la durée anticipée du trajet en voiture.

  Si un prédicteur crédible annonce publiquement 8h de bouchon, les gens renoncent tous à prendre la voiture et la prédiction sera réfutée.

  Si annonce trafic fluide, tout le monde prend la voiture, prédiction réfutée.

  ->Une prédiction publique correcte doit être compatible avec son annonce. Donc dans un système où le prédicteur public est fiable, le trafic est toujours « modéré ».

-> Les équilibres d'un système social ne sont pas les mêmes selon qu'on y fait ou non usage d'un outil de prédiction.

Deux points :

1. La réflexivité n'est pas incompatible avec la fiabilité (le modèle de prédiction peut intégrer les informations pertinentes sur les conditions de la diffusion de ses oracles et les conséquences sociales des anticipations produites par cette diffusion)
2. Dans un certain nombre de cas la réflexivité *renforce* la prédictibilité

3.5 Prédictions, anticipations, coordination

- Phénomènes sociaux : agrégats de comportements individuels
- Comportements individuels : déterminés par des préférences et anticipations

-> en particulier : anticipation des comportements d'autrui

- Anticipation = prédiction

Mais multiplicité des équilibres : plusieurs anticipations également « rationnelles »

-> prédiction difficile

L'existence d'une prédiction publique crédible contribue à la co-sélection d'un équilibre (« l'avenir c'est le format pdf »)

-> La prédiction publique est un outil de coordination et donne lieu à des prophéties auto-réalisatrices (SFP)

3.5 La crédibilité ou quand prédire c'est choisir un peu

Exemple : un prédicteur réputé peut annoncer avant le 1er tour de 2007 :

1. « Le président élu sera Bayrou ou Sarkozy ». Donne aux sympathisants de Ségolène une raison de voter Bayrou. Qui passe au second tour. Bayrou ou Sarkozy est élu. -> prédiction valide
2. « Le président élu sera Ségolène ou Sarkozy ». Donne aux sympathisants de Bayrou une raison de voter SR ou NS, qui passent au second tour. ->prédiction valide

-> Condition nécessaire de ce scenario auto-réalisateur :

(*) le prédicteur doit avoir *une crédibilité et une visibilité* suffisante.

Mais (*) a d'autant moins de chance d'être réalisée que nous sommes mauvais prédicteurs (cf. plus haut), qu'il y a concurrence des prédictions (Marx sur l'avenir du capitalisme) etc.

-> Quid avec le progrès futur des IA prédictives en sciences sociales ?

3.6 Réflexivité + IA

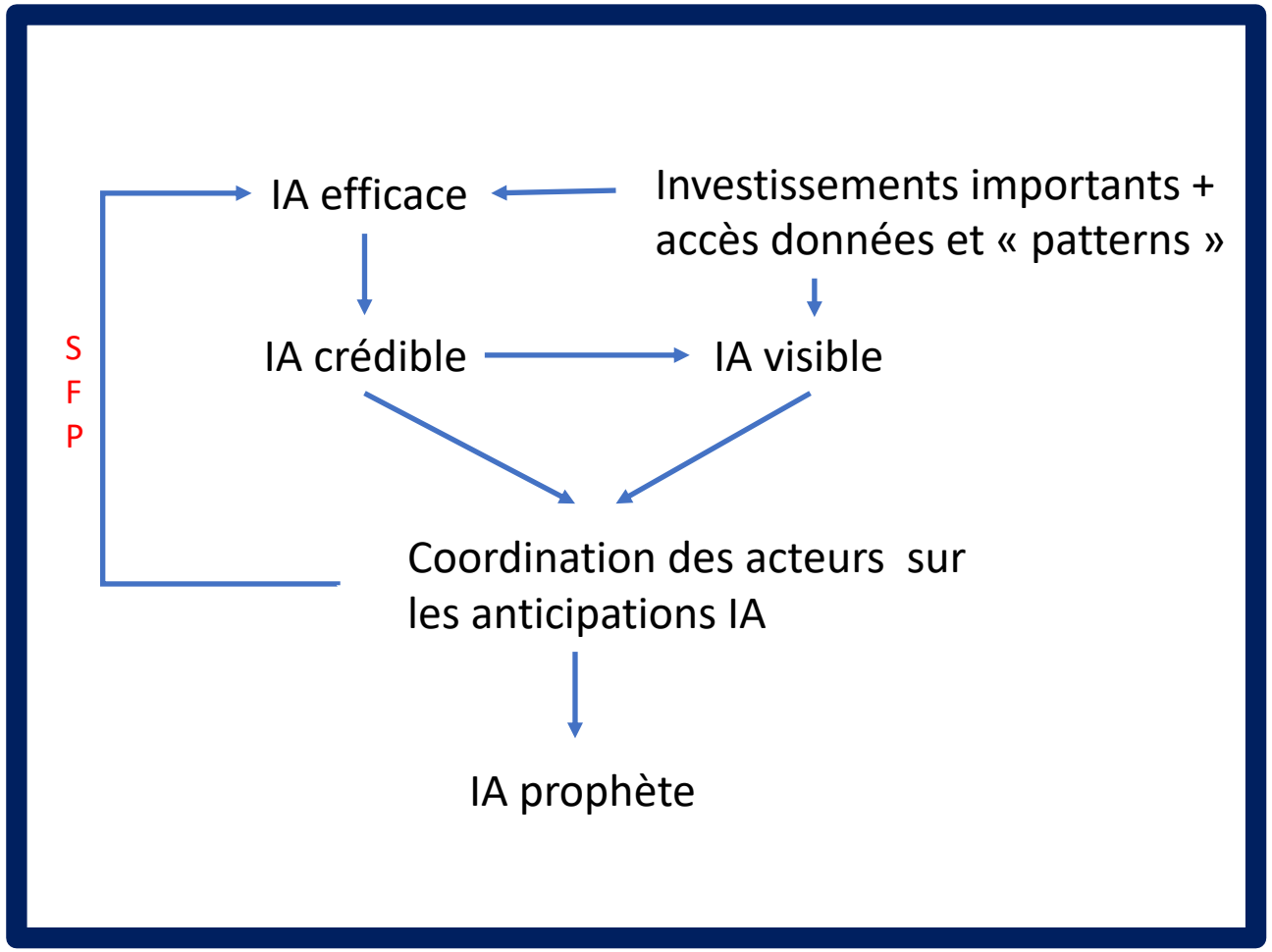
- Supp. une IA sociale a un historique prédictif long et parfait et qu'il soit vrai et évident pour tous que l'intérêt bien compris des propriétaires soit de garantir sa fiabilité. Il pourrait devenir ordinaire et rationnel de fonder son action sur la base de ses prédictions sociales (météo sociale).
- L'IA peut être fiable et programmée pour annoncer (1) ou (2) ; ou peut *par accident* annoncer l'un ou l'autre indifféremment.

Deux problèmes :

1. Transparence de l'IA : comment détecter la manipulation/le pb si les algo utilisés ne sont pas accessibles (privés, protégés par un droit étranger etc.) ou pas interprétables ?
2. Visibilité de l'IA (pouvoir des gafam) : même si la résolution tendancieuse de l'indétermination de l'avenir est détectée, l'issue non-désirée s'impose s'il n'existe aucun point alternatif de coordination sur une issue meilleure.

3.7 Silicone prophète

Bank run (Merton)	Retirer mes dépôts	Laisser mes dépôts
Retirer mes dépôts	(0, 0) E.N. risque-dominant	(0, -2)
Laisser mes dépôts	(-2, 0)	(1, 1) E.N. Pareto-optimal



Quelle raison pour anticiper que les autres vont laisser leur dépôt ?
-> enjeu de crédibilité des anticipations : crédibilité + visibilité = pouvoir de coordination

3.8 Responsabilités

- La prédiction publique crédible est une action stratégique qui engage l'avenir collectif

  (+ contexte risqué : prévisible défaillance des marchés de service de prédiction IA du fait des coûts d'entrée, de l'accès aux données etc.)

-> justifie l'exercice d'un contrôle public sur l'activité

- Puissance publique : Insee, Commissariat au Plan, France Stratégie, BCE, …

vs.

Emergence de puissances digitales déterritorialisées

(+ contexte : séparatisme individualiste /« décitoyennisation »/défiances démocratiques)

= mouvement d'aliénation des outils de coordination (transfert de la puissance de coordination hors du contrôle démocratique)

Références

Sur les prophéties auto-réalisatrices et les sciences sociales :

- Barnes B., « Social life as bootstrapped induction », Sociology 17 524-545
- Biggs M,. Self-fulfilling prophecies, in Handbook of analytical sociology, chap. 13., Oxford University Press, 2011
- Dupuy J.P., *L'avenir de l'économie*, Flammarion 2012
- Giddens A., *Central Problems in Social Theory*, MacMillan,1979
- Henshel R.L., « The Boundary of the Self-Fulfilling Prophecy and the Dilemma of Social Prediction », *The British Journal of Sociology*, 33 (4) pp. 511-528, 1982.
- Lagespertz, *The Opposite Mirrors*, Kluwer, 1995
- MacKenzie, *An Engine, Not a Camera. How financial models shape markets*, MIT press, 2006
- Merton R.K., « The unanticipated consequences of purposive social action », American sociological review 1 894-904, 1936
- Merton R.K. « The Self-fulfilling Prophecy » (1948)
- Salganik M. and Watts D., « An experimental approach to Self-fulfilling Prophecies in cultural markets », 2007
- Walliser, Anticipations, équilibres et rationalité économiques, Claman-Levy, 1985

Prédictibilité en sciences sociales et « computational social sciences » :

- Gigerenzer G., « Mindless statistics », The journal of socio-economics 33, 2004
- Open Science collaboration, « Estimating the reproducibility of psychological science », Science 349, 2015
- Salganik M, Dodds P.S., Watts D., « Experimental study of inequality and unpredictablity in an artificial cultural market », Science 311, 2006
- Martin T, Hofman J, Sharma A, Anderson A, Watts D, « Exploring limits to prediction in complex social systems », WWW'16, 2016
- Hoffman, Sharma, Watts, « Prediction and explanation in social systems », Science 355, 2017 (n° spécial prédiction)
- Lazer & al., Computational social science, Science 323 2009
- Conte & al. , Manifesto of computational social science, European Physical journal 214, 2012

Annexe : critique de pratique statistique en sciences sociales
« Mindless model fitting vs prediction »

« Fitting a model to data that is already obtained is not sound hypothesis testing, even if the resulting explained variance, or R2, is impressive. The reason is that one does not know how much noise one has fitted, and the more adjustable parameters one has, the more noise one can fit. Psychologists habitually fit rather than predict, and rarely test a model on new data, such as by cross-validation (Roberts and Pashler, 2000). Fitting per se has the same problems as story telling after the fact, which leads to a “hindsight bias” (Hoffrage et al., 2000). The true test of a model is to fix its parameters on one sample, and to test it in a new sample. Then it turns out that predictions based on simple heuristics can be more accurate than routine multiple regressions (Czerlinski et al., 1999). Less can be more. The routine use of linear multiple regression exemplifies another mindless use of statistics. » (Gigerenzer 2004)