Étude et développement d'un modèle simple de Reinforcement Learning

Alexis Emanuelli, sous la supervision de Stefano Vrizzi

ÉCOLE NORMALE SUPÉRIEURE | DÉPARTEMENT D'ÉTUDES COGNITIVES

LNC²

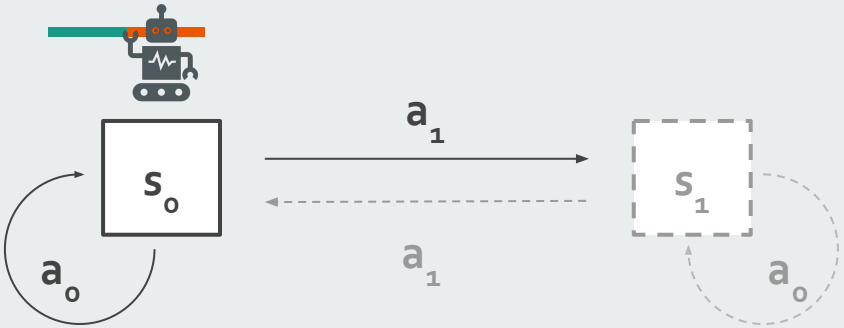
Introduction

1. Familiarisation : Rescorla Wagner pour un agent
2. Modéliser plusieurs agents qui s'influencent
3. Reproduction de données empiriques à l'aide du modèle.

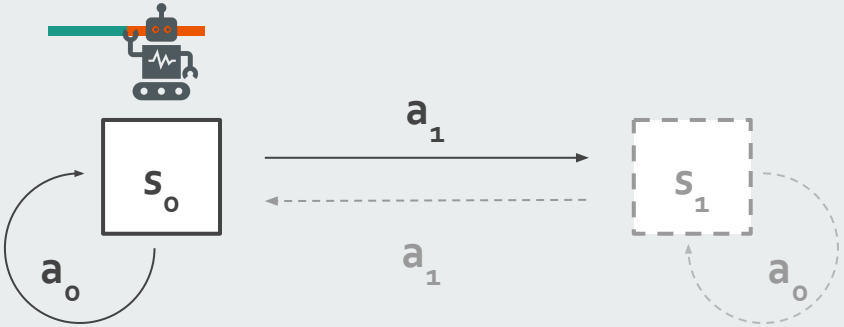
Méthodes : anticiper les résultats des graphiques : discussion précédent toute implémentation

Familiarisation : Rescorla Wagner pour un agent

Familiarisation : Rescorla Wagner pour un agent



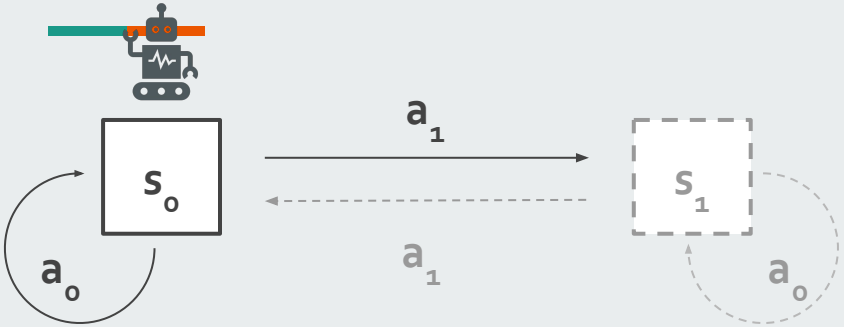
Familiarisation : Rescorla Wagner pour un agent



Q = valeur attendue d'une action dans un état du modèle

	a_0	a_1
s_0	Q_{00}	Q_{01}
s_1	Q_{10}	Q_{11}

Familiarisation : Rescorla Wagner pour un agent



Q = valeur attendue d'une action dans un état du modèle

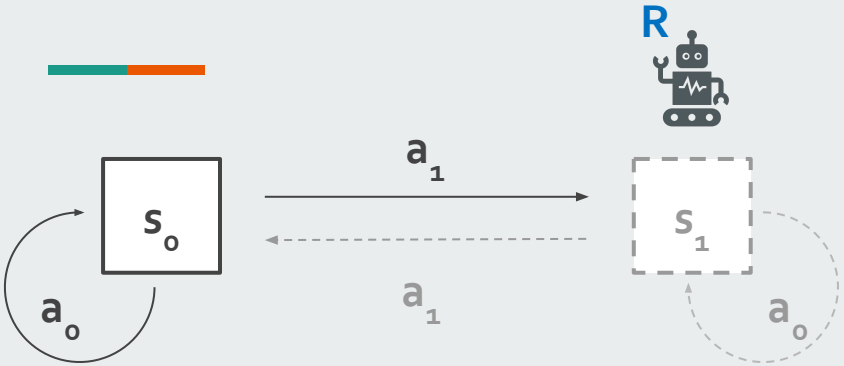
	a_0	a_1
s_0	Q_{00}	Q_{01}
s_1	Q_{10}	Q_{11}

$$p(s,a^*) = \frac{e^{Q(s,a^*)\beta}}{\sum_a e^{Q(a,s)\beta}}$$

Softmax Decision

e.g. a_1

Familiarisation : Rescorla Wagner pour un agent



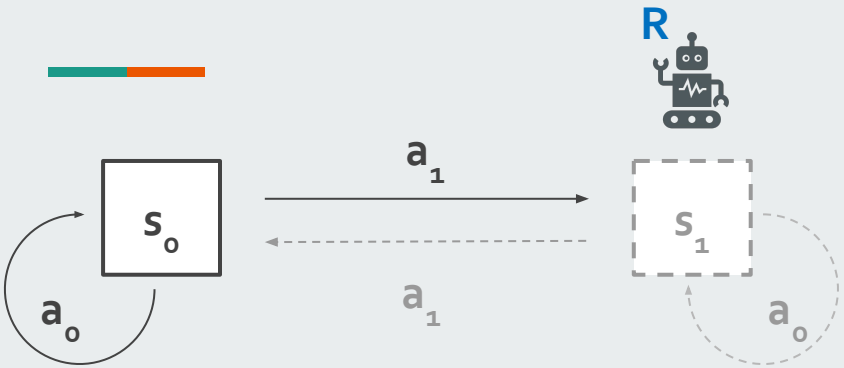
	a_0	a_1
s_0	Q_{00}	Q_{01}
s_1	Q_{10}	Q_{11}

BASIC MODEL

$$Q_{t+1} = Q_t + \alpha\, \delta(R-Q_t)$$

1) Mesurer la surprise :
différence entre le résultat R et la valeur supposée Q

Familiarisation : Rescorla Wagner pour un agent



	a_0	a_1
s_0	Q_{00}	Q_{01}
s_1	Q_{10}	Q_{11}

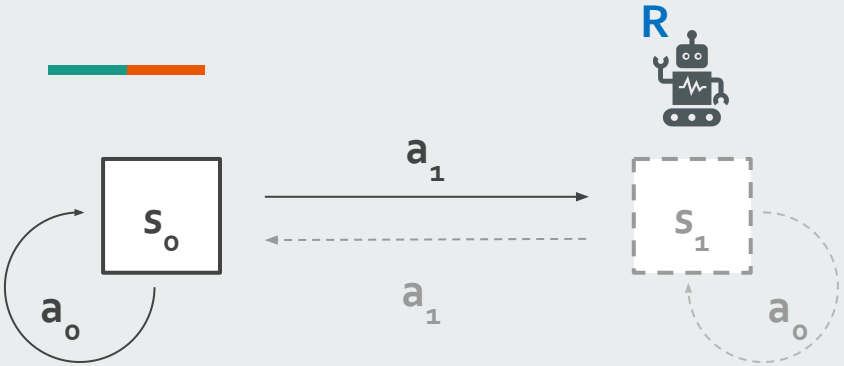
BASIC MODEL

$$Q_{t+1} = Q_t + \alpha\, \delta(R - Q_t)$$

1) Mesurer la surprise :
différence entre le résultat R et la valeur supposée Q)

2) L'échelonner par le paramètre **α** (entre 0 et 1)

Familiarisation : Rescorla Wagner pour un agent



	a_0	a_1
s_0	Q_{00}	Q_{01}
s_1	Q_{10}	Q_{11}

BASIC MODEL

$$Q_{t+1} = Q_t + \alpha\, \delta(R - Q_t)$$

1) Mesurer la surprise :
   différence entre le résultat R et la valeur supposée Q)

2) L'échelonner par le paramètre **α** (entre 0 et 1)

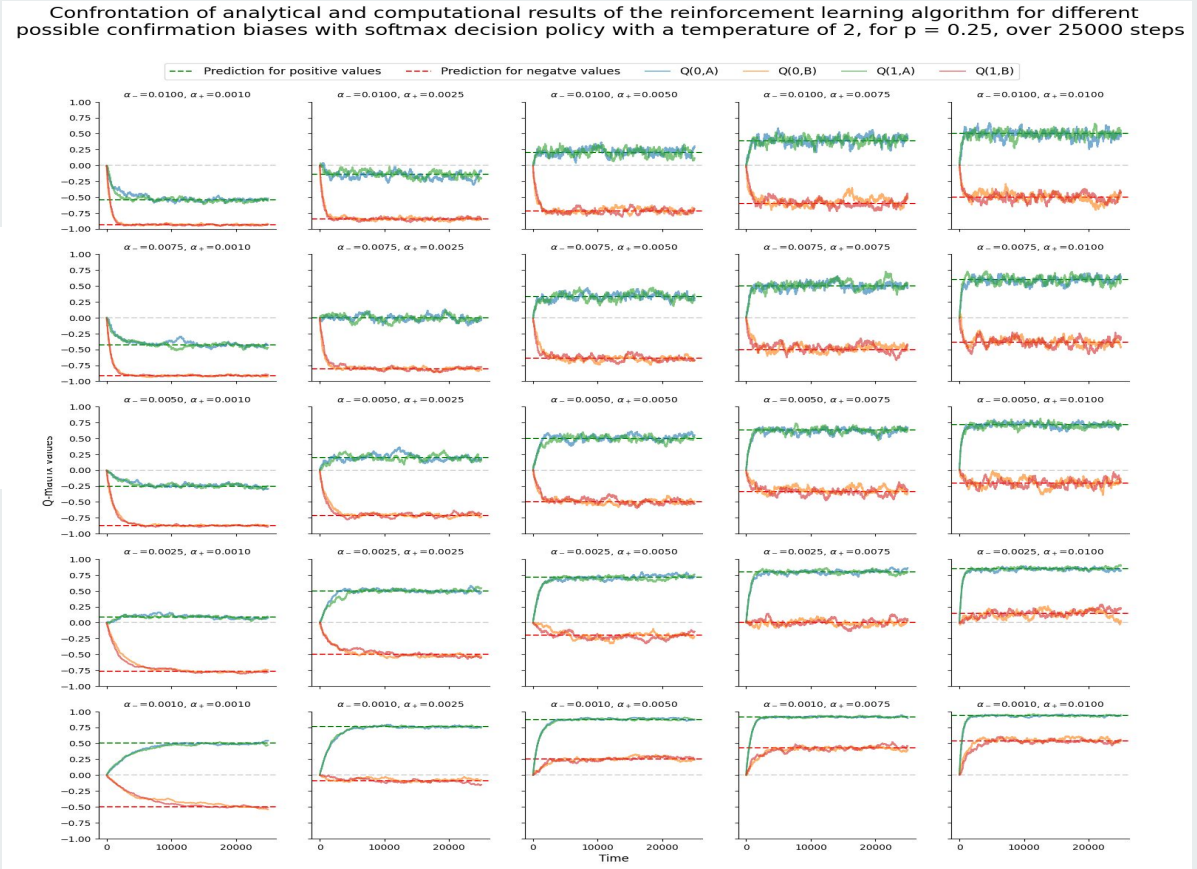
3) Mise à jour de Q

Ajout du “confirmation bias”

learning rates

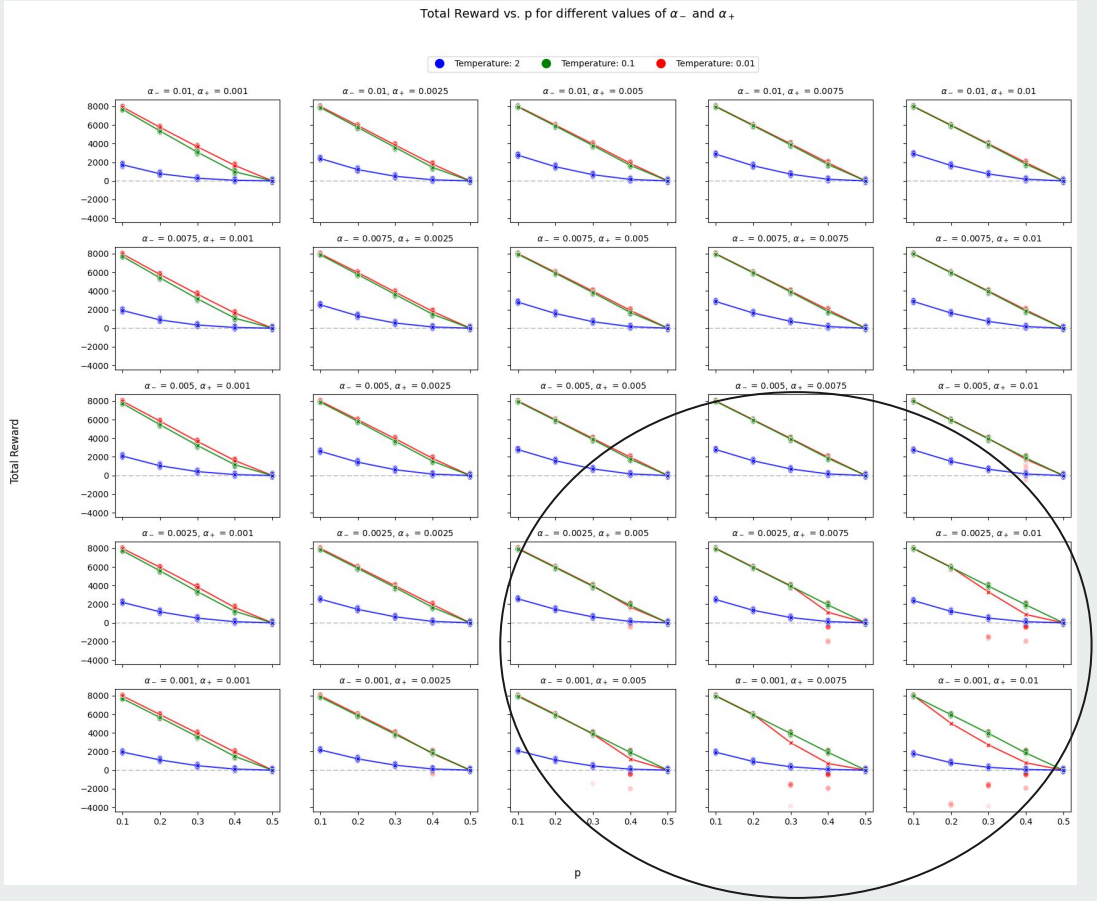
$$y_{\text{minus}} = \frac{p \cdot \alpha_{\text{plus}} - (1-p) \cdot \alpha_{\text{minus}}}{p \cdot \alpha_{\text{plus}} + (1-p) \cdot \alpha_{\text{minus}}}$$

$$y_{\text{plus}} = \frac{(1-p) \cdot \alpha_{\text{plus}} - p \cdot \alpha_{\text{minus}}}{(1-p) \cdot \alpha_{\text{plus}} + p \cdot \alpha_{\text{minus}}}$$

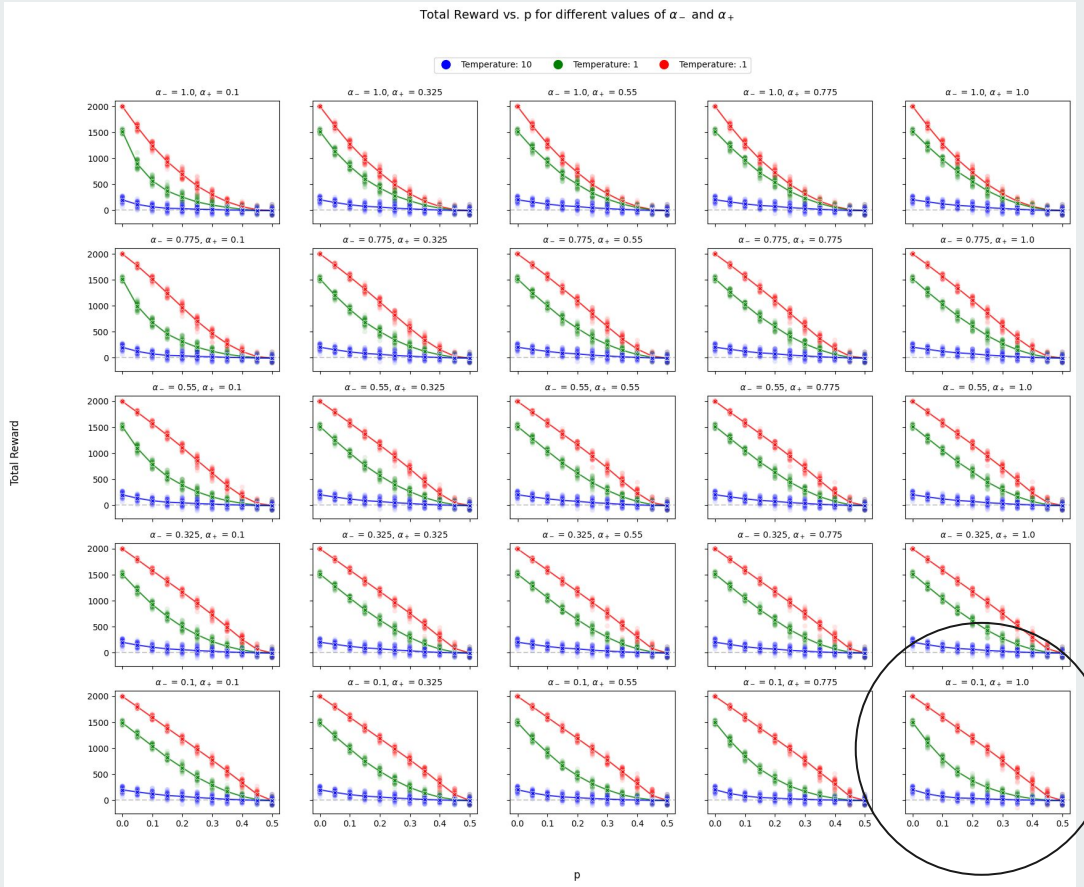


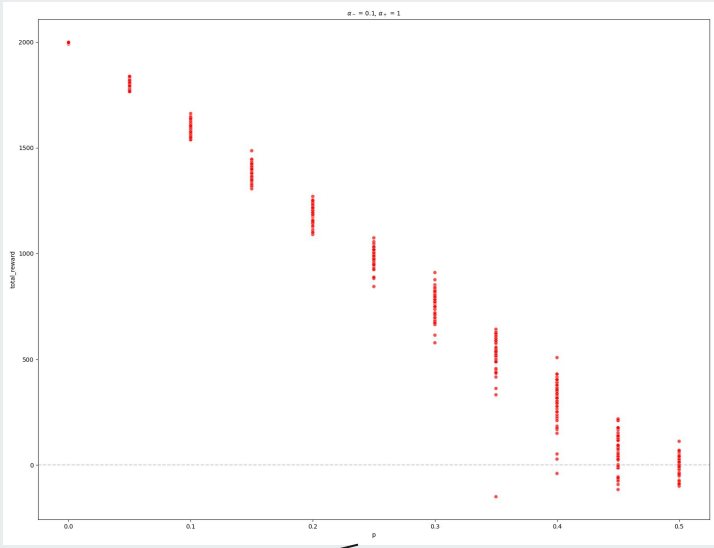
Comment devenir aveugle à ses propres erreurs

Learning rate moyens & faible temperature



Comment devenir aveugle à ses propres erreurs





Très grands learning rates et plus hautes températures

Reproduction de résultats empiriques

Adapter le modèle à un cas réel

REPORTS

Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market

Matthew J. Salganik,[1,2]* Peter Sheridan Dodds,[2]* Duncan J. Watts[1,2,3]*

Hit songs, books, and movies are many times more successful than average, suggesting that "the best" alternatives are qualitatively different from "the rest"; yet experts routinely fail to predict which products will succeed. We investigated this paradox experimentally, by creating an artificial "music market" in which 14,341 participants downloaded previously unknown songs either with or without knowledge of previous participants' choices. Increasing the strength of social influence increased both inequality and unpredictability of success. Success was also only partly determined by quality: The best songs rarely did poorly, and the worst rarely did well, but any other result was possible.

Adapter le modèle à un cas réel

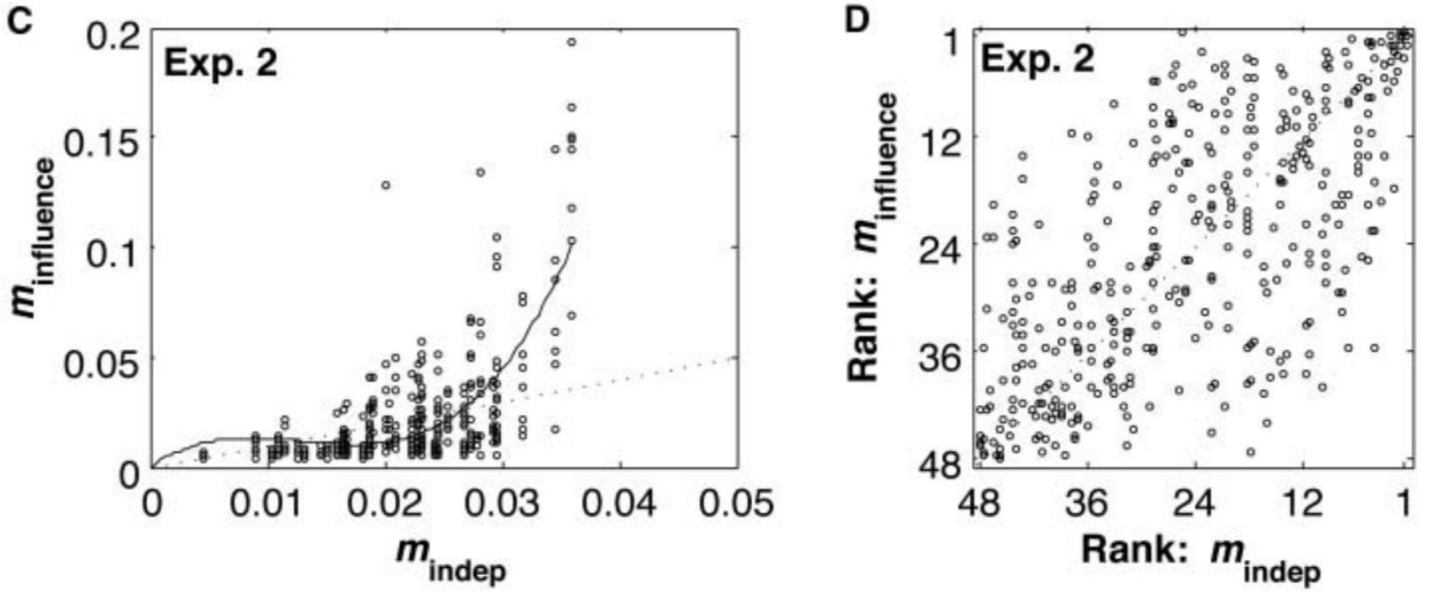


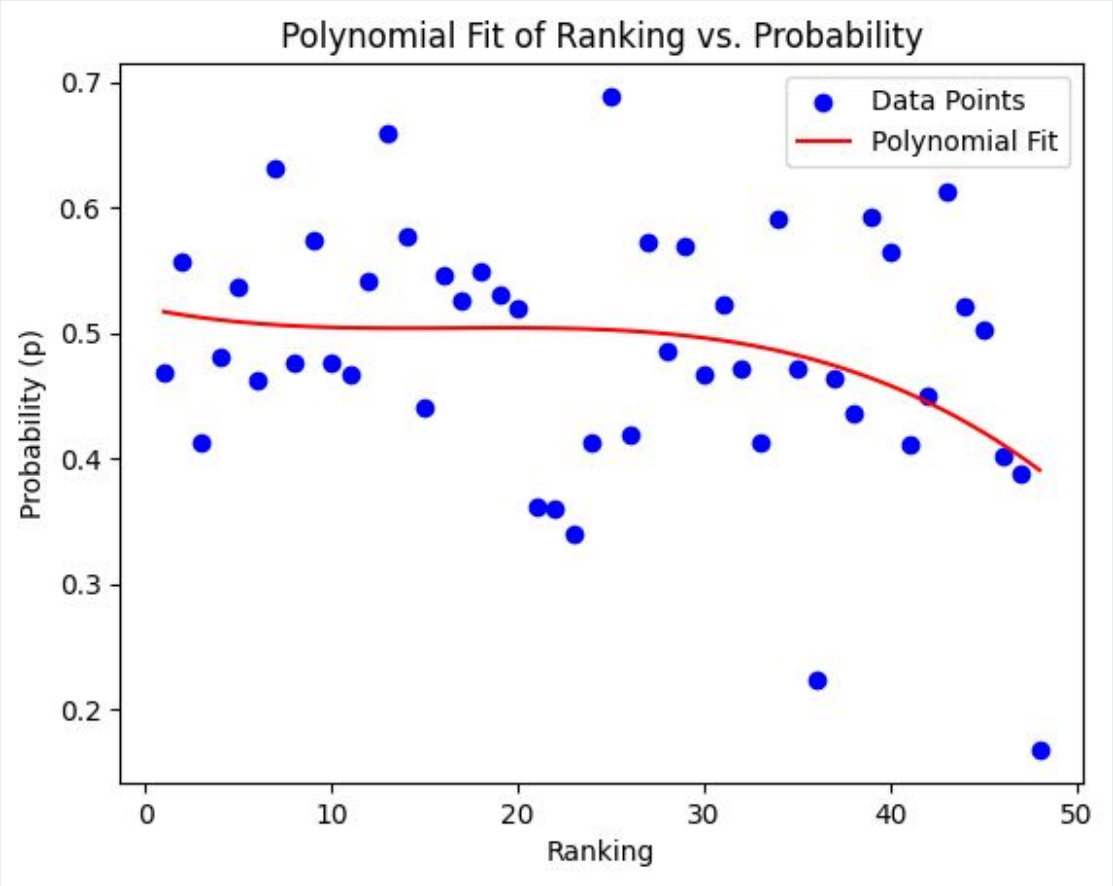
Fig. 3. Relationship between quality and success. (**A**) and (**C**) show the relationship between m_{indep}, the market share in the one independent world (i.e., quality), and $m_{influence}$, the market share in the eight social influence worlds (i.e., success). The dotted lines correspond to quality equaling success. The solid lines are third-degree polynomial fits to the data, which suggest that the relationship between quality and success has greater convexity in experiment 2 than in experiment 1. (**B**) and (**D**) present the corresponding market rank data.

Adapter le modèle à un cas réel

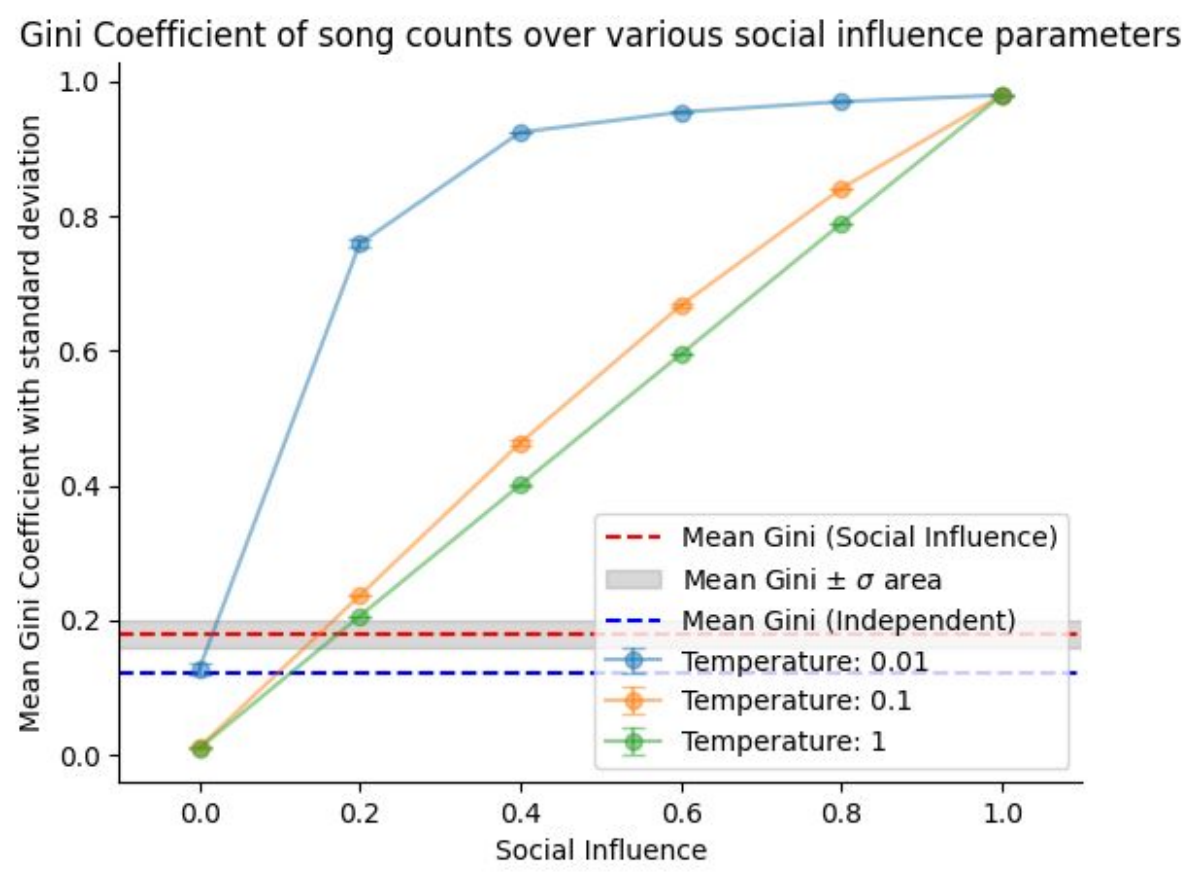
1 seul état dans lequel chaque action correspond à une chanson.

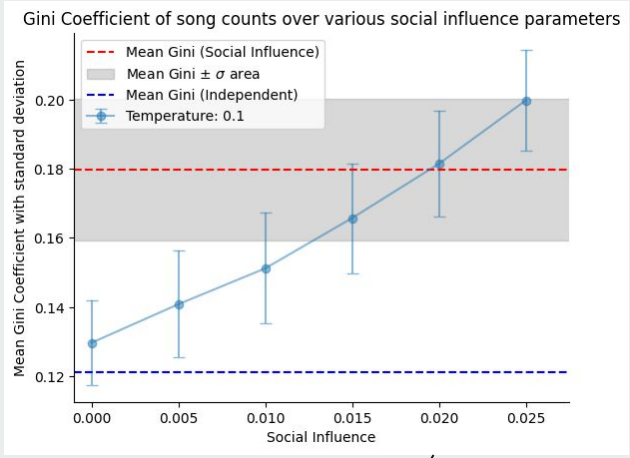
- Comment choisir les probas de récompense : la qualité intrinsèque des chansons ?
  —> fit polynomial (degré3) des données expérimentales, en fonction du classement
- Comment modéliser l'influence sociale ?
  —> deuxième tirage aléatoire entre la chanson la plus écoutée jusqu'ici et la chanson choisie par l'agent avec le softmax
- Comment choisir les paramètres du modèle ? —> épouser au mieux le gini index des données empiriques

Adapter le modèle à un cas réel



Adapter le modèle à un cas réel





Social influence = 0.02

Adapter le modèle à un cas réel