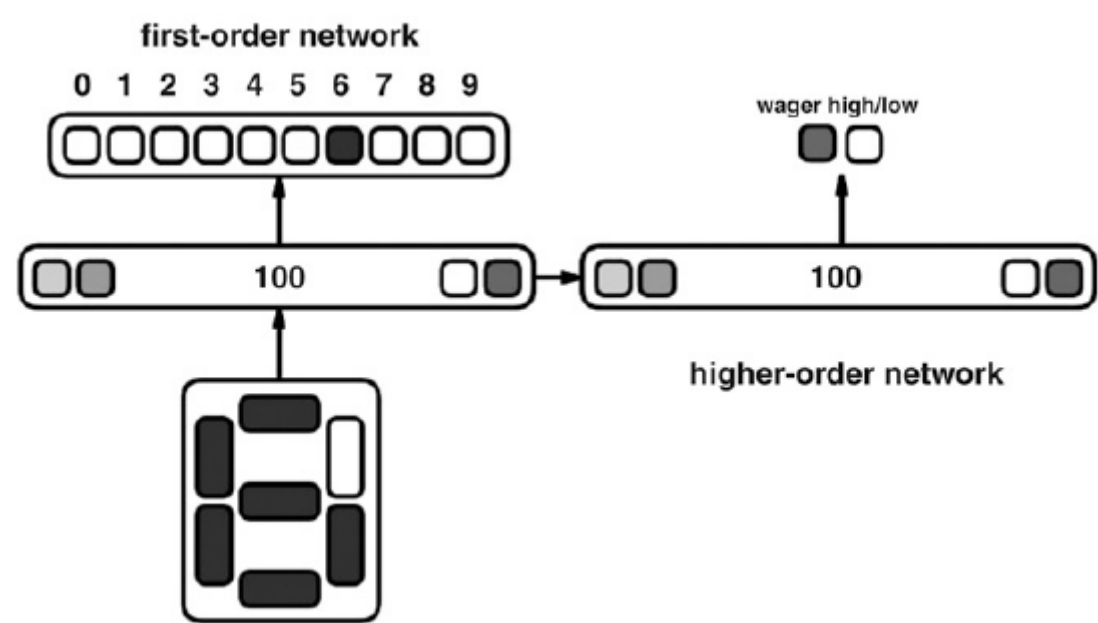


Consciousness and metarepresentation : A computational sketch

[Axel Cleeremans, Bert Timmermans, Antoine Pasquali]

Simulation 2 : Reconnaissance de chiffre digitaux



Principe

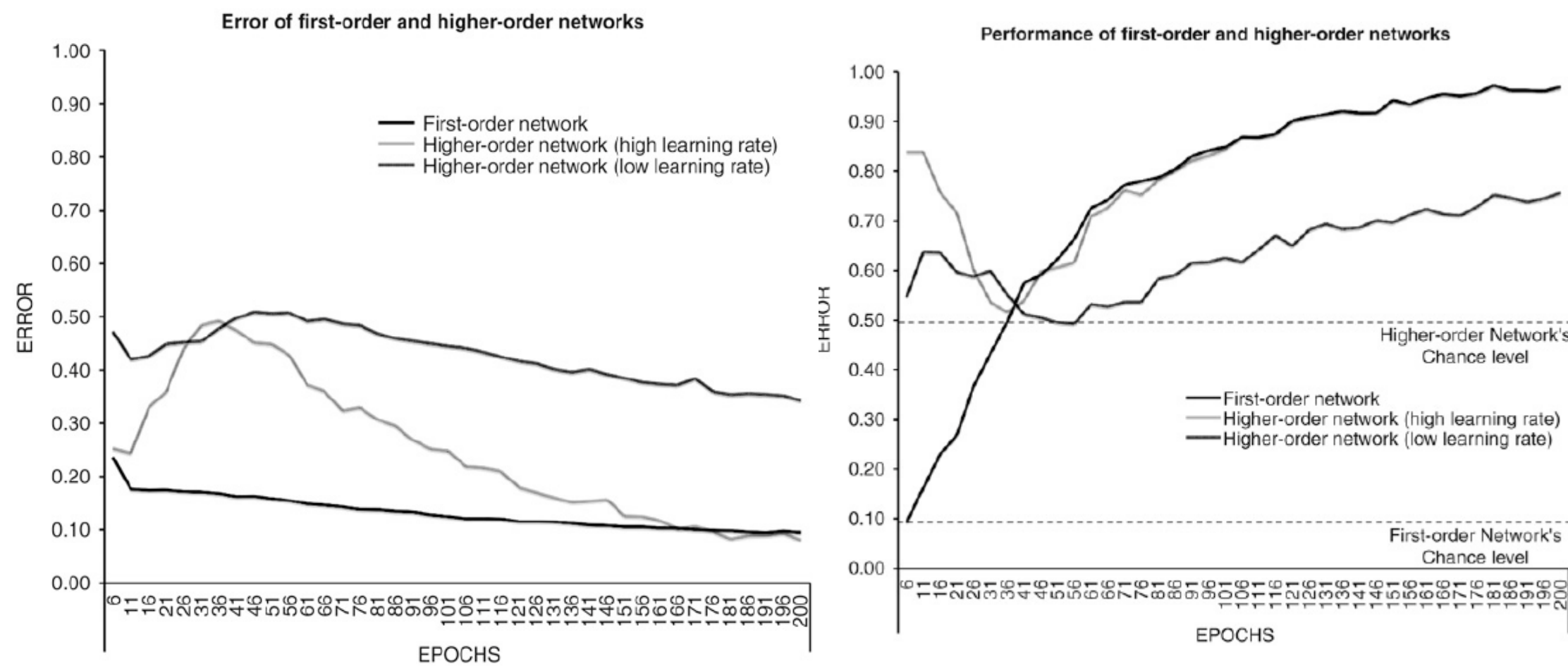
Un premier réseau apprend à associer une forme à un chiffre. En même temps, un second réseau apprend à parier sur le résultat du premier réseau, à partir, seulement, de la couche caché du premier.

Noter que lors de l'apprentissage du second réseau, il a aussi accès aux sorties du premier et à celles sortie désirées.

Paramètres

Momentum : 0.5 pour le premier réseau	Taux d'apprentissage : 0.15 pour le premier réseau	Moyenne sur 10 réseaux
Apprentissage : 200 epochs	0.1/ 10 ⁻⁷ taux d'apprentissage (constant) pour les réseaux sup.	100 unités cachés pour tous

Résultat de l'article



Conclusion

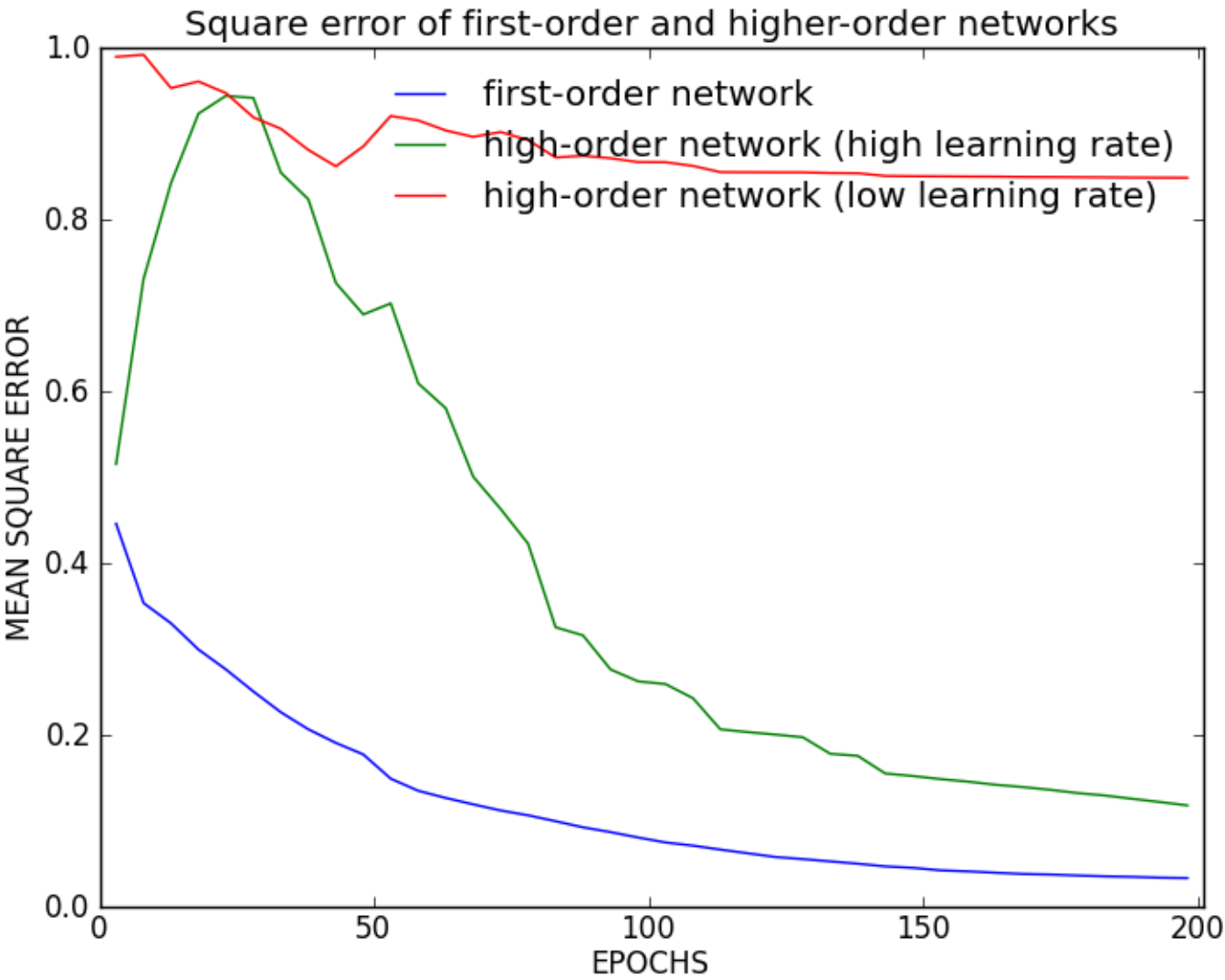
Les réseaux d'ordres supérieurs se trompent et dé-apprennent durant les premières epochs et commencent seulement à se stabiliser ensuite.

Ce qui est du au fait qu'au départ, les réseaux de second ordre paris continuellement sur l'échec, sauf que le premier réseau commence à réussir à reconnaître les formes. (voir bonus paris confiants)

Notons que les réseaux d'ordre supérieur n'arrivent pas à dépasser les performances du premier après les bonnes reconnaissances. (le comparateur y remédiera)

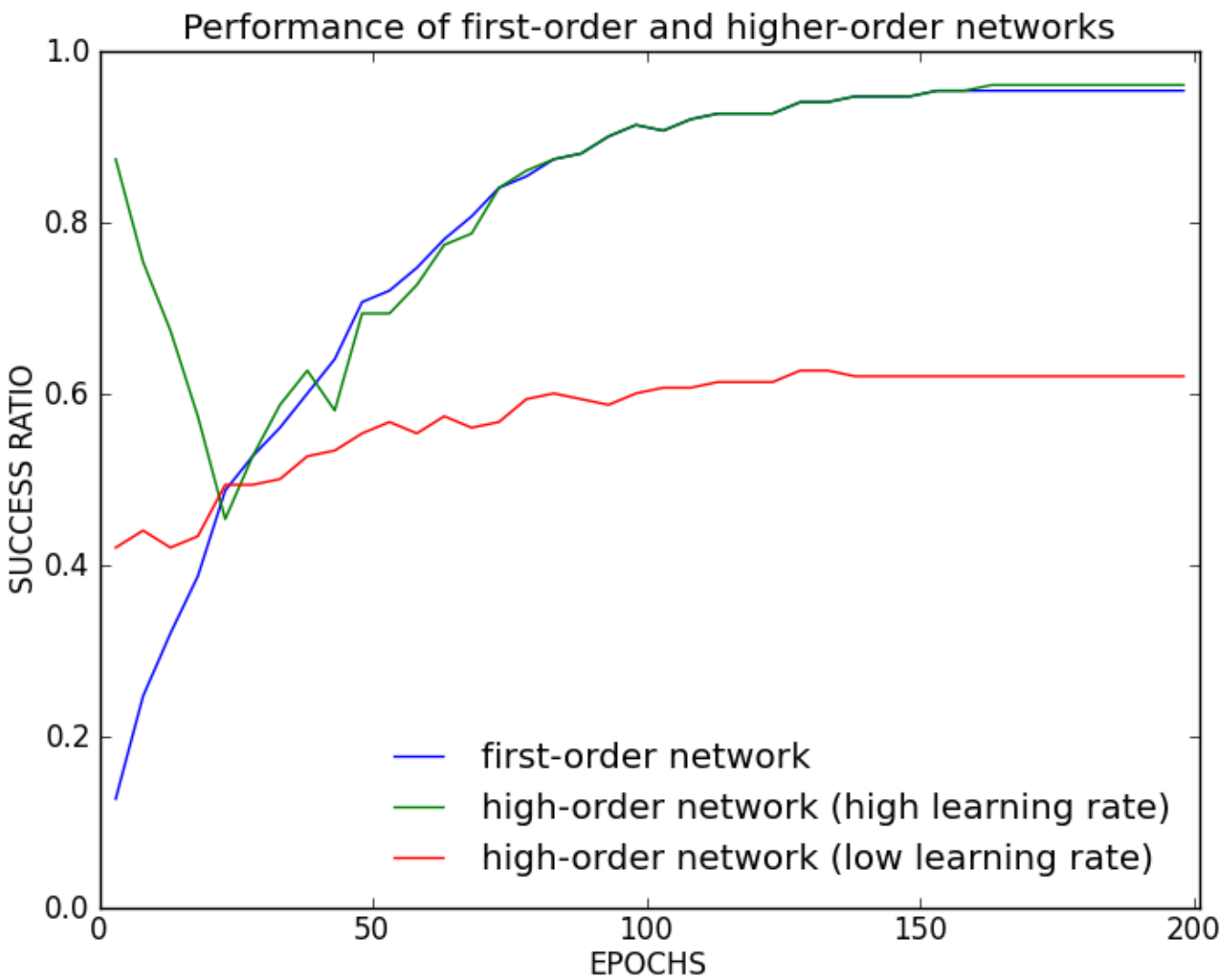
Nos résultats :

Paramètres devinés/supposés			
Température : 1 pour tous	Taux d'apprentissage constant	Rétro propagation online	Momentum : 0. pour les supérieurs
Sorties des neurones sur [0 ; 1]	Sigmoïde sur [0,1] : $\frac{1}{1+e^{-\theta x}}$	Apprentissage : 200 epochs x 10 formes (shuffle)	Poids initialisés entre -1 et 1 pour le premier réseaux
Poids initialisés entre -0.25 et 0.25 pour les seconds réseaux	Biais activés		



N'ayant pu identifier précisément la courbe décrite par l'article, nous dessinons l'erreur moyenne au carrée

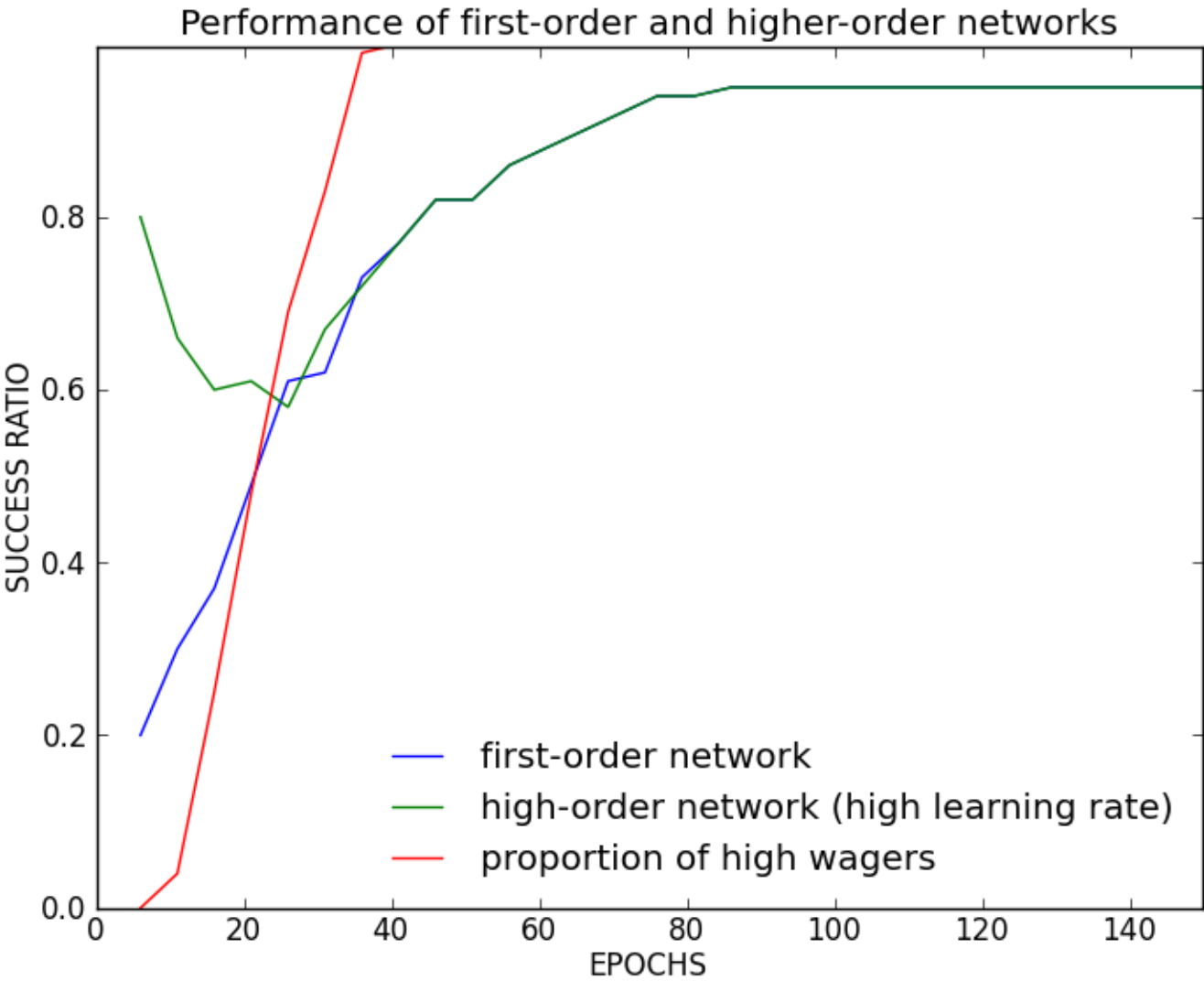
Les tendances sont les mêmes.



Les performances sont similaires.

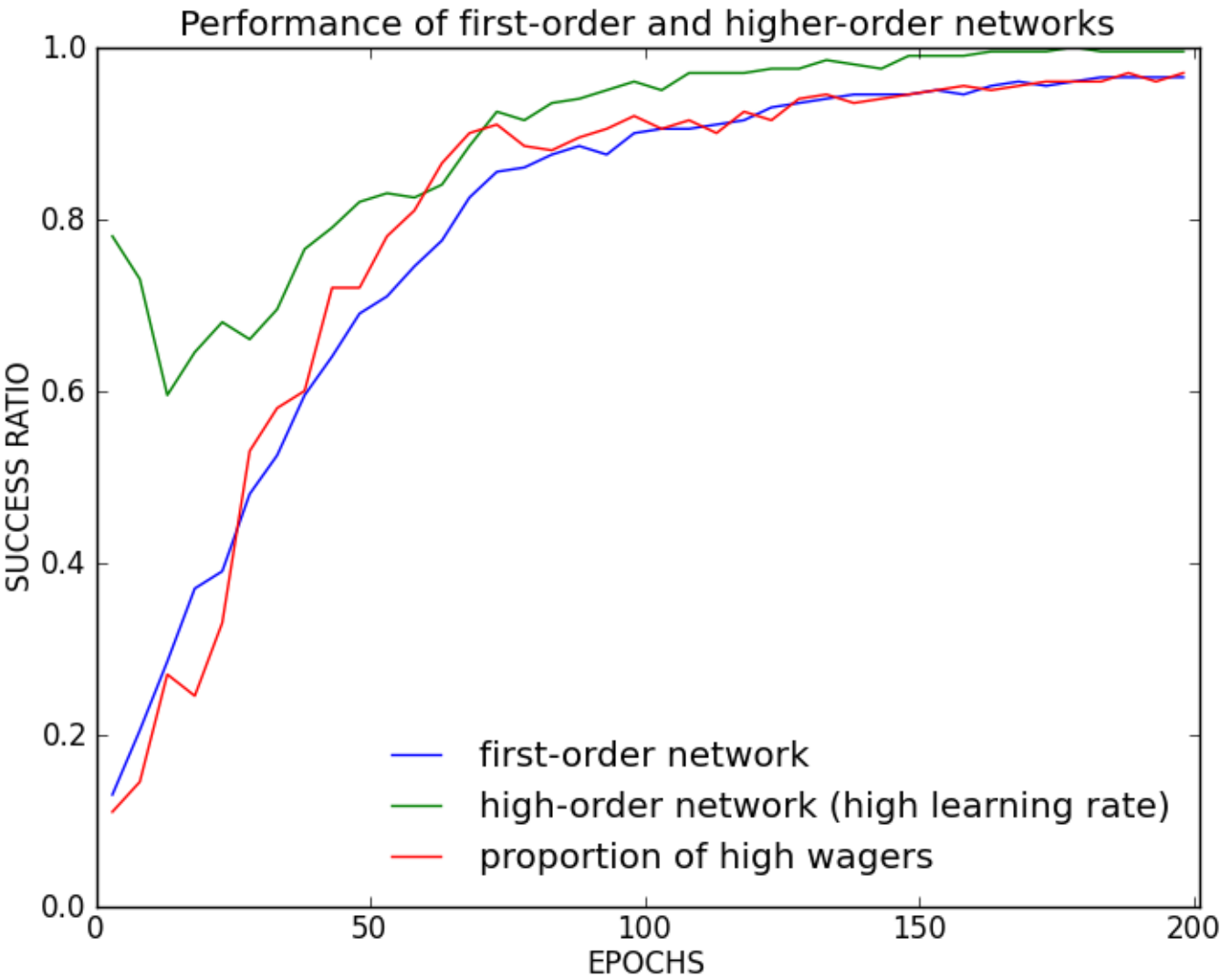
On remarque que la courbe du réseau supérieur ne dépasse pas celle du premier réseau, ce qui signifie qu'il ne pari pas assez intelligemment.

Voir la courbe suivante pour plus de précision.

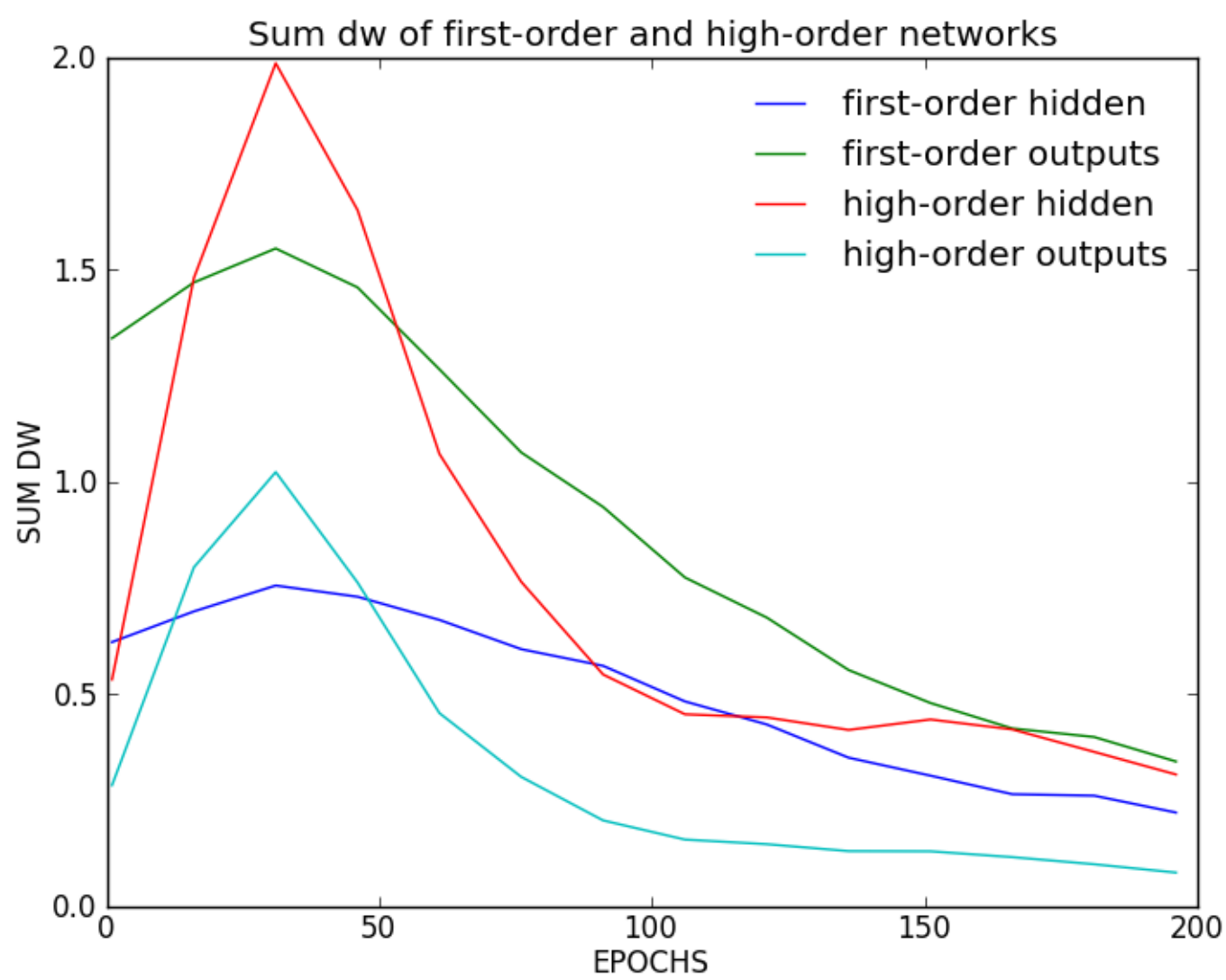
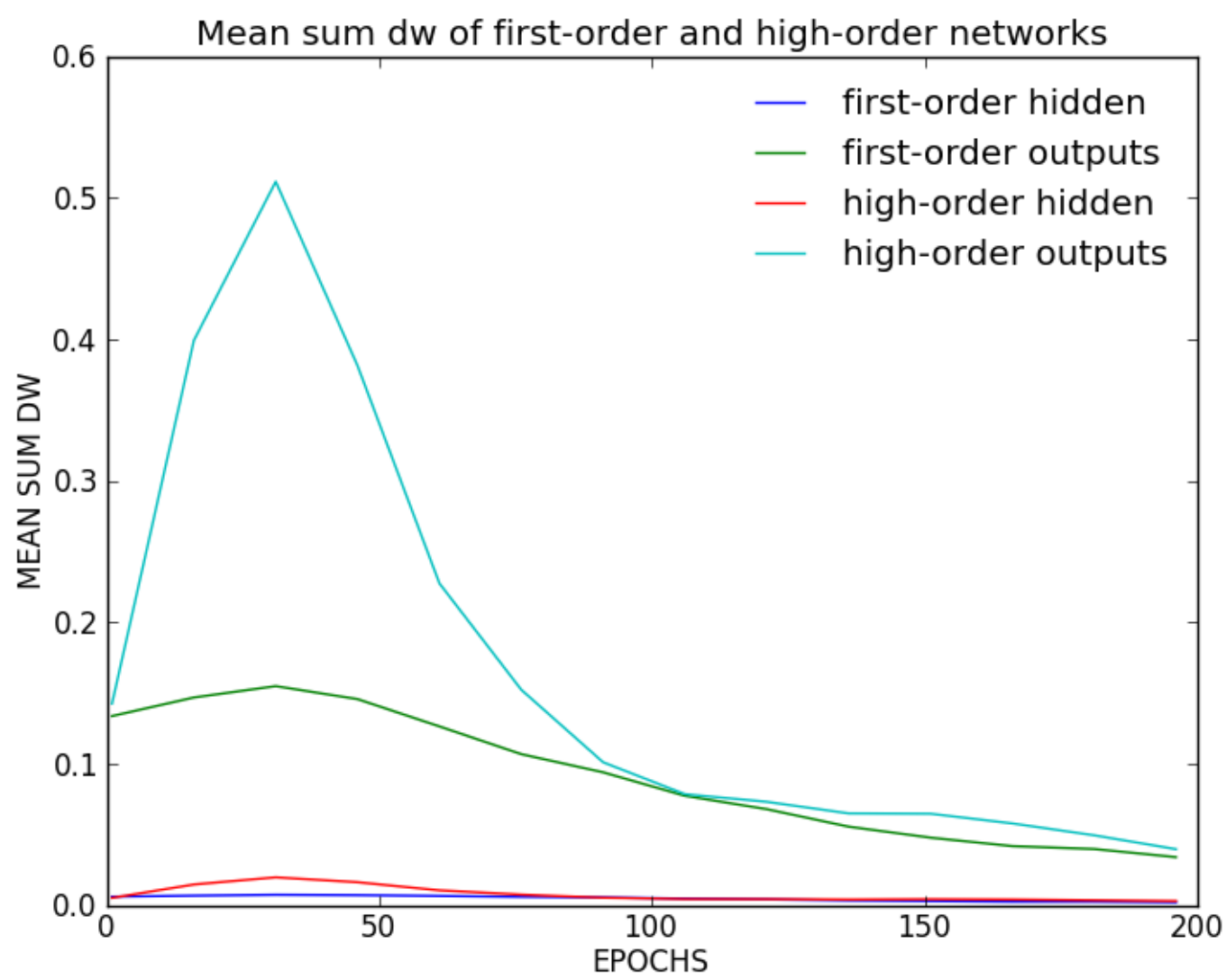


Le second réseau apprend simplement à parier haut en même temps que le premier reconnaît mieux les formes.

Néanmoins, il n'arrive pas à le dépasser et se contente de parier haut en continu.



Notons tout de même qu'en initialisant les poids du second réseaux sur $[-1 ; 1]$, avec un momentum de 0.5 (comme le premier réseau) : le réseau d'ordre supérieur est alors capable de parier bien plus efficacement.



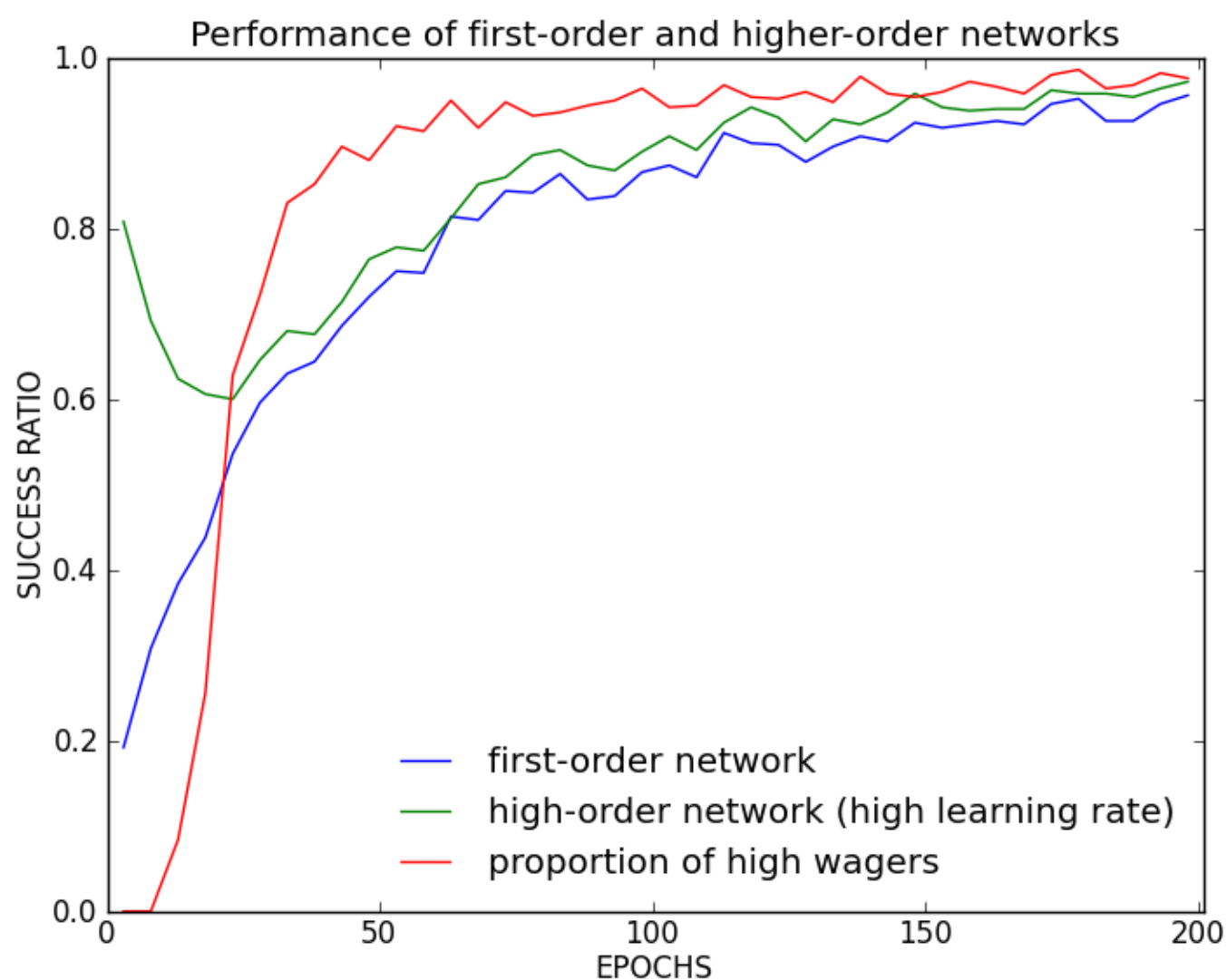
Essais sur des chiffres manuscrits

Entrées : 16 x 16 pixels

100 unités caché pour le premier

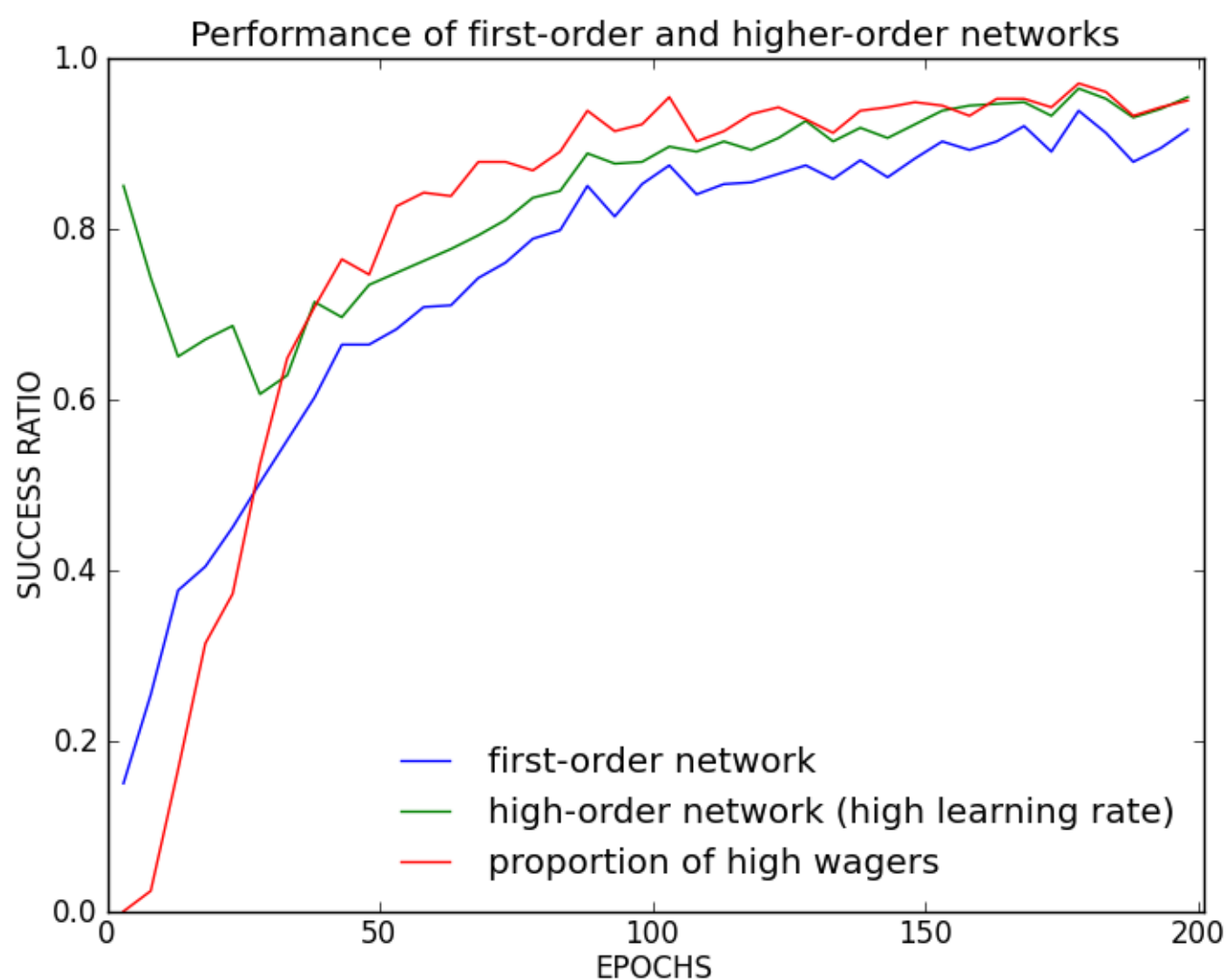
20 unités caché pour le second

Apprentissage : 200 epochs x 50 formes (shuffle) avec 1600 formes



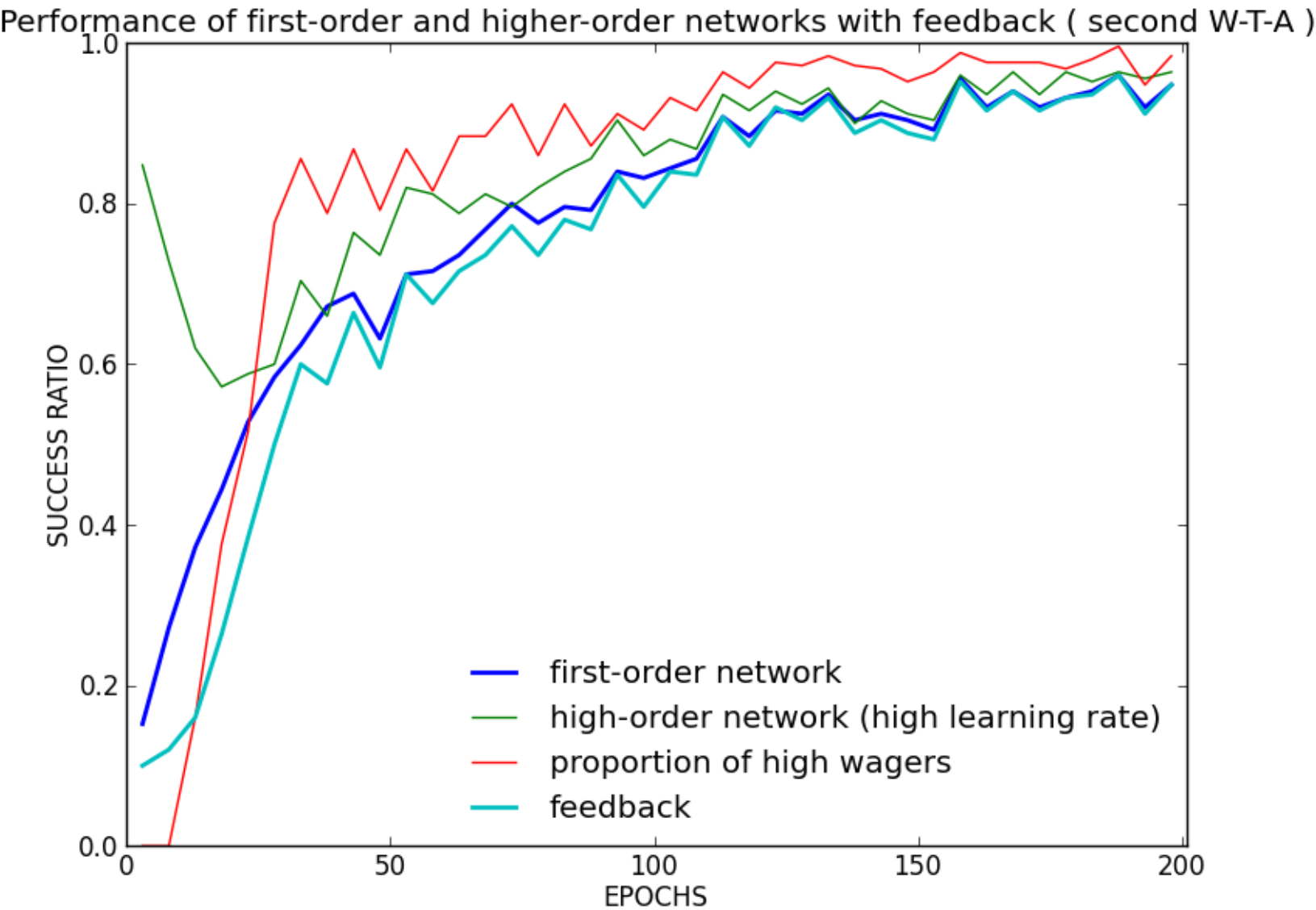
Les courbes sont très similaires à l'expérience précédente.

Bonne nouvelle, les performances ne sont pas détériorées sur de vrais chiffres.



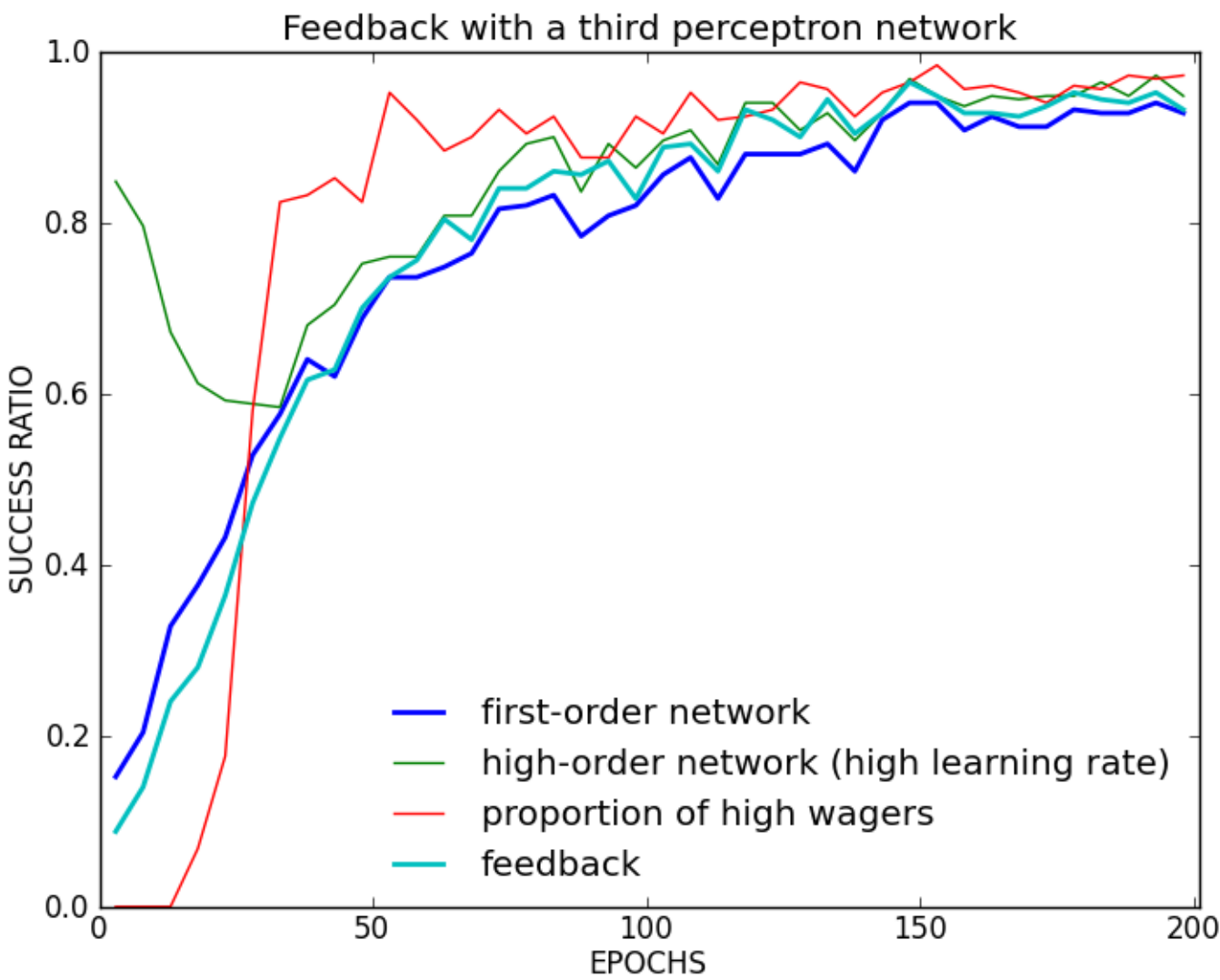
Nous avons refait l'essai en initialisant le second réseaux avec les mêmes paramètres que le premier.

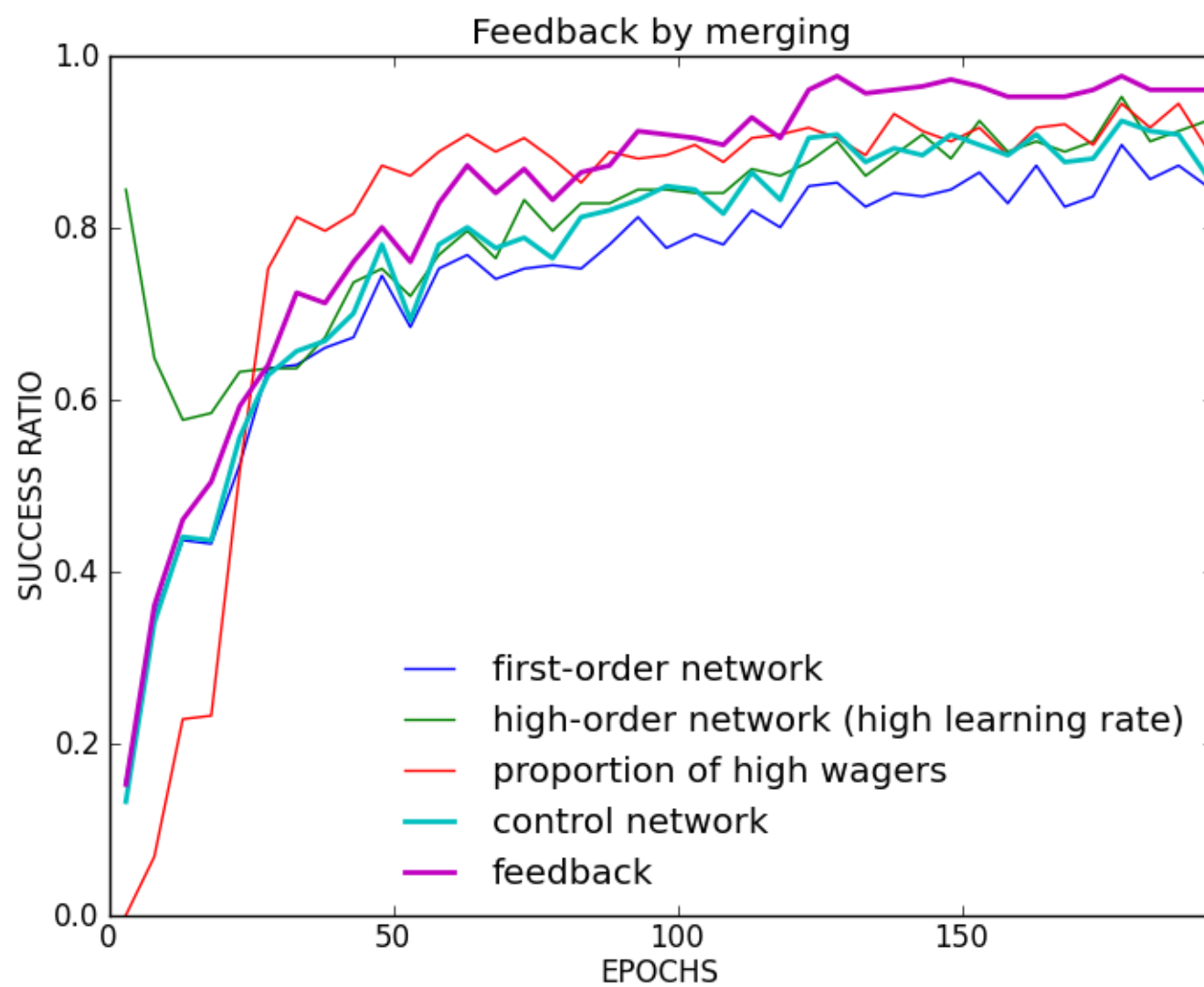
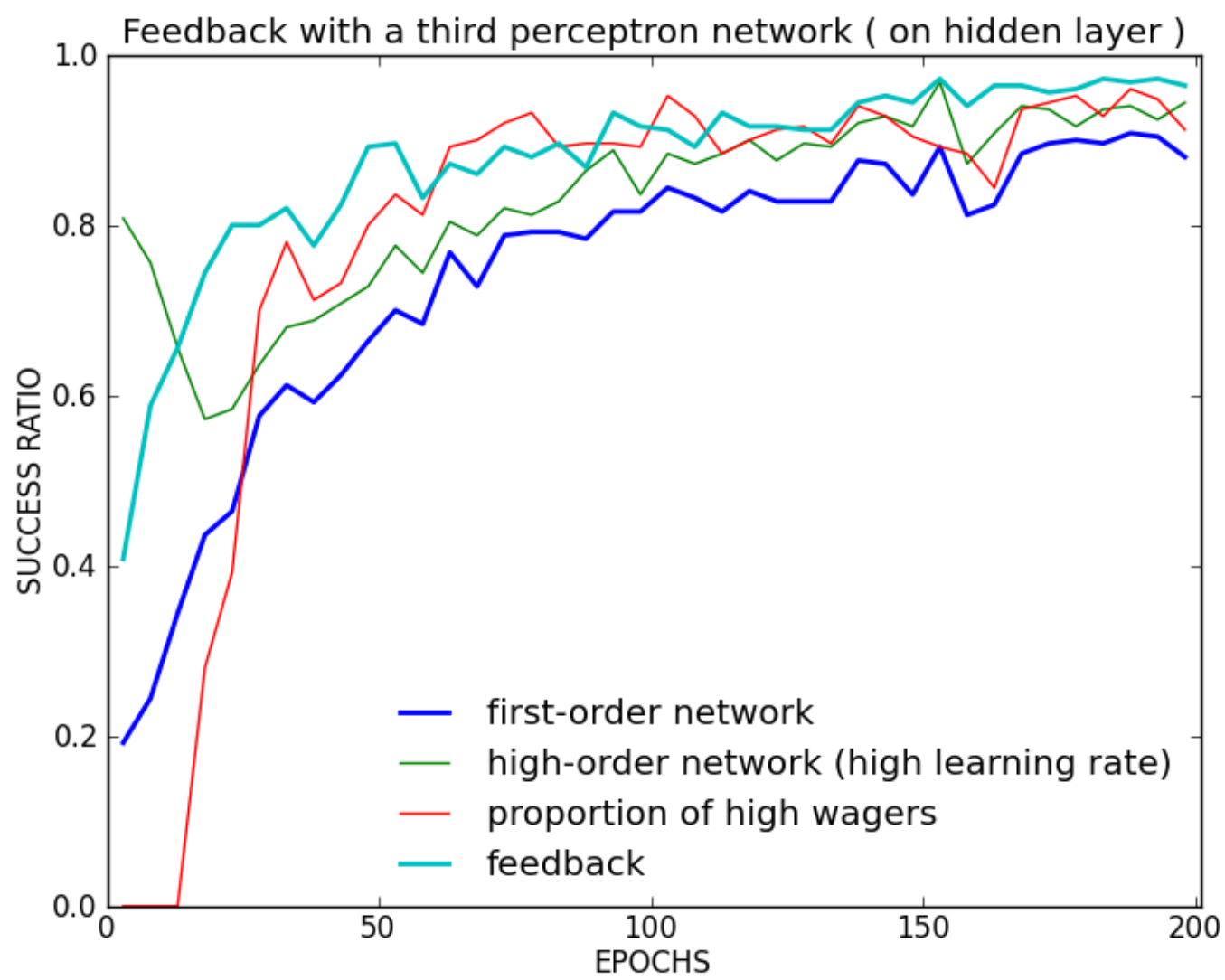
Nous nous apercevons cette fois-ci que les performances ne sont pas autant améliorées que sur la tâche « facile ».



Les performances ne sont pas améliorées pour plusieurs raisons :

- la répartition de la réponse dans le nième neurone le plus actif est la suivante :
1 : 0.3 | 2 : 0.12 | 3 : 0.12 | 4 : 0.1 | 5:0.8 | 6 : 0.1 | 7 : 0.1 | 8 : 0.1 | 9 : 0.01
- le taux de paris hauts est trop élevé, donc le comportement avec ou sans feedback est le même (les performances du réseaux de second ordre ne dépassent pas assez celles du premier)





On connecte les 2 sorties du second réseau comme entrée de la couche de sortie du premier réseau.

La courbe bleu est le calcul du premier réseau sans le feedback (18 entrées mis à 0)

Les 2 neurones de sorties sont amplifiées sur 18 entrées (pour être significatif par rapport aux 100 de la couche caché).

La mise à jour des poids du feedback suit approximativement une règle d'apprentissage de perceptron.

Contrairement à ce qu'on voit, les performances gagnées grâce au feedback ne sont pas si évidentes car il faut bien prendre en compte que l'apprentissage du premier réseau est dégradé par les 18 entrées mises à 0.

Performance of first-order and higher-order networks with feedback (nth W-T-A

