

Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

Critique par Félix Morissette et Rémi Thériault
PSY 7101 : Métho
Université du Québec à Montréal

Crédibilité de l'article

Titre	Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect	Titre attrayant	Titre spécifique
Journal	Journal of Personality and Social Psychology (JPSP)	Facteur d'impact : 5.919	Revue la plus prestigieuse en psychologie sociale
Auteur	Daryl J. Bem Professeur Émérite en psychologie à Cornell University, éditeur associé de JPSP	Créateur de la théorie de l'autoperception en psychologie sociale	Auteur de plus de 150 articles, cités près de 30 000 fois, l'un des psychologues sociaux les plus connus et respectés en psychologie

Why Are Scientists Afraid of Daryl Bem?

"When one village dog barks, the rest take up the howl."

—Zen saying

"If I want to stop a research program I can always do it by getting a few experts to sit in on the subject, because they know right away that it was a fool thing to try in the first place."¹

—Charles F. Kettering, Head of research, General Motors, 1920-1947

Prejudice: from the Latin, "judgment in advance"²

D prejudice is never far from the experience of most readers of *Explore*,



Daryl J. Bem, PhD

entists have become hostile toward what they consider Republicans' contempt for the basics of modern biology, anthropology, evolutionary theory, geology, cosmology, stem cell research, and climate change.⁴

The Pew findings imply that scientists are liberal, progressive thinkers who are tolerant of new ideas. Unfortunately, this is not always the case.

BEM'S BOMBSHELL

Flagrant prejudice among scientists erupted in early 2011, when Cornell University psychology professor Daryl Bem had a paper accepted for publication in the elite *Journal of Personality and Social Psy-*



By Daniel Engber

Daryl Bem Proved ESP Is Real

Which means science is broken.

MAY 17, 2017 • COVER STORY

Social Psychology Should Retract

Why the Journal of Personality and Social Psychology Should Retract Article DOI:

10.1037/a0021524

“Feeling the Future:

Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. Two variants of psi are precognition (*knowing* (conscious cognitive awareness) and *pronostication* (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process. Precognition and pronostication are themselves special cases of a more general phenomenon: the anomalous retroactive influence of some future event on an individual's current responses, whether those responses are conscious or unconscious, cognitive or affective. This article reports 9 experiments, involving more than 1,000 participants, that test for retroactive influence by "time reversing" well-established psychological effects so that the individual's responses are obtained before the putatively causal stimulus events occur. Data are presented for 4 time-reversed effects: precognitive approach to erotic stimuli and precognitive avoidance of negative stimuli; retroactive priming; retroactive habituation; and retroactive facilitation of recall. The mean effect size (d) in psi performance across all 9 experiments was .21, and all but one of the experiments yielded statistically significant results. The individual-difference variable of stimulus seeking, a component of extraversion, was significantly correlated with psi performance in 5 of the experiments, with participants who scored above the midpoint on a scale of stimulus seeking achieving a mean effect size of 0.43. Skepticism about issues of replication, and theories of psi are also discussed.

The psychology of parapsychology, or why good researchers publishing good articles in good journals can still get it totally wrong

Unless you've been pleasantly napping under a rock for the last couple of months, there's a good chance you've heard about a forthcoming article in the Journal of Personality and Social Psychology (JPSP) purporting to provide strong evidence for the

Construits

- Psi (VD) :
 - Télépathie
 - Clairvoyance/vision à distance
 - Psychokinésie
 - Précognition (✓)
 - Prémonition (✓)



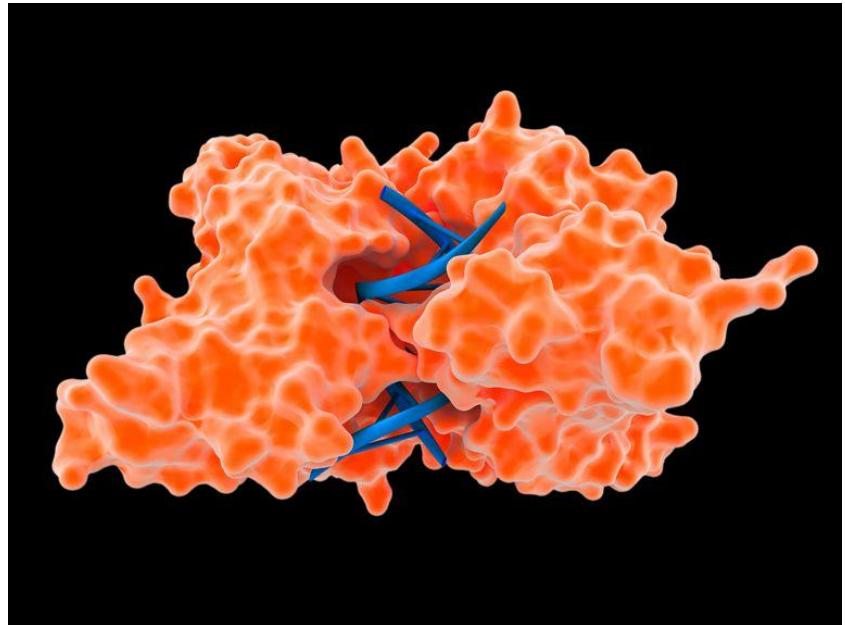
Contexte théorique

Théorie évolutionniste :

- augmente les chances de reproduction et aide à éviter les menaces et autres dangers

Mécanisme de fonctionnement :

La physique quantique (?)



Contexte théorique : Critique

Physique quantique : comprends pas

Psi et conscience : comprends pas

Donc physique quantique = psi et conscience

$$1 + 1 = 1$$

Mise en contexte

Question de recherche :

Est-ce possible d'influencer de manière rétroactive la cognition et les affects? En d'autres mots : est-ce que le futur peut influencer le présent?

Hypothèse générale :

Les réponses cognitives et affectives des participants seront affectées par des stimuli auxquels ils n'auront *pas encore été exposés*... (la flèche du temps sera inversée)

Participants

- Étudiant.es universitaires non-gradué.es recruté.es dans le département de psychologie de l'université Cornell.
- Minimum de 100 participant.es par étude, pour assurer une taille d'effet
- « Most effect sizes (d) reported in the psi literature range between 0.2 and 0.3. If $d = 0.25$ and $N = 100$, the power to detect an effect significant at .05 by a one-tail, one-sample t test is .80 (Cohen, 1988). » (p.3)
- Les études comptaient respectivement : 100, 150, 100, 100, 150, 200, 100, et 50 étudiant.es.
- Taux d'attrition nul/non rapporté.
- Recrutement à l'aide d'un logiciel en ligne de recrutement de participant.es.

Participant.es et critique

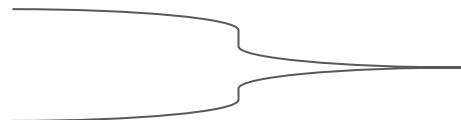
- Échantillonnage de commodité
- Biais de sélection : attirer ceux qui pensent avoir des pouvoirs psychiques
- Taille d'échantillon variable (raisons non expliquée)

9 études(!)

1) Détection précognitive de stimuli érotiques

2) Évitement précognitif des stimuli négatifs

3) Amorçage rétroactif I



Amorçage (priming) affectif

4) Amorçage rétroactif II



Habituation

5) Habituation rétroactive I



6) Habituation rétroactive II

7) Induction rétroactive d'un sentiment d'ennui

8) Facilitation rétroactive du rappel I

9) Facilitation rétroactive du rappel II

Facilitation du rappel

Méthodologie générale

Passation des tests

1. Lecture d'un court texte d'introduction
2. Trois minutes de relaxation
3. Tâche d'une vingtaine de minutes
4. Les participant.es reçoivent 5\$ ou un crédit en psychologie expérimentale

Texte de recrutement et d'introduction

« This is an experiment that tests for ESP. It takes about 20 minutes and is run completely by computer. First you will answer a couple of brief questions. Then, on each trial of the experiment, pictures of two curtains will appear on the screen side by side. One of them has a picture behind it; the other has a blank wall behind it. Your task is to click on the curtain that you feel has the picture behind it. The curtain will then open, permitting you to see if you selected the correct curtain. There will be 36 trials in all. Several of the pictures contain explicit erotic images (e.g., couples engaged in nonviolent but explicit consensual sexual acts). If you object to seeing such images, you should not participate in this experiment. »

1) Détection précognitive de stimuli érotiques

Tâches et mesures :

- 1) Images tirées du International Affective Picture System (IAPS; Lang & Greenwald, 1993)
- 2) 36 essais/participant.es
- 3) **La réussite pour trouver des images érotiques VS non-érotiques**
- 4) **L'hypothèse de départ était que les participants allaient identifier la présence d'image érotique significativement plus que la chance.**
- 5) Possibilité de choisir des sets d'image homo ou hétérosexuelle, et modification des images pour hommes

From: Daryl J. Bem <djb5@cornell.edu>

Sent: November 8, 2019 1:19 PM

To: Rémi Thériault <remi.theriault@mail.mcgill.ca>

Subject: Re: Demo of precognition task available?

Hi,

I'm not aware of any demos of my precognition tasks currently online. However, now that I have retired from Cornell, I no longer have a physical laboratory, so I have shifted to running my experiments on the web, where anyone can take part. I am planning to post a version of my "Retroactive Facilitation of Recall (Experiments 8 & 9 in my article) and run it as a full experiment.

I have to finish some other tasks before I can do that, but I will try to get it up online before the end of the current semester and they will receive a brief summary of their performance. I will keep you posted and give you the link to the experiment when it is ready to go. You wouldn't have to try to reproduce the experiment since it will be self-contained. And if your students list your University when they sign on, I can send you the full data for them to you. Also, it shouldn't matter whether or not they have been told the details of the experiment ahead of time (including the quote about the White Queen that introduces Experiment 8 in my article).

Best.

Daryl J. Bem

d.bem@cornell.edu







1) Détection précognitive de stimuli érotiques

Hypothèse : Les participants seront capable d'identifier la position future de la photo érotique significativement plus que 50% du temps car les gens sont motivés à détecter les stimuli érotiques.

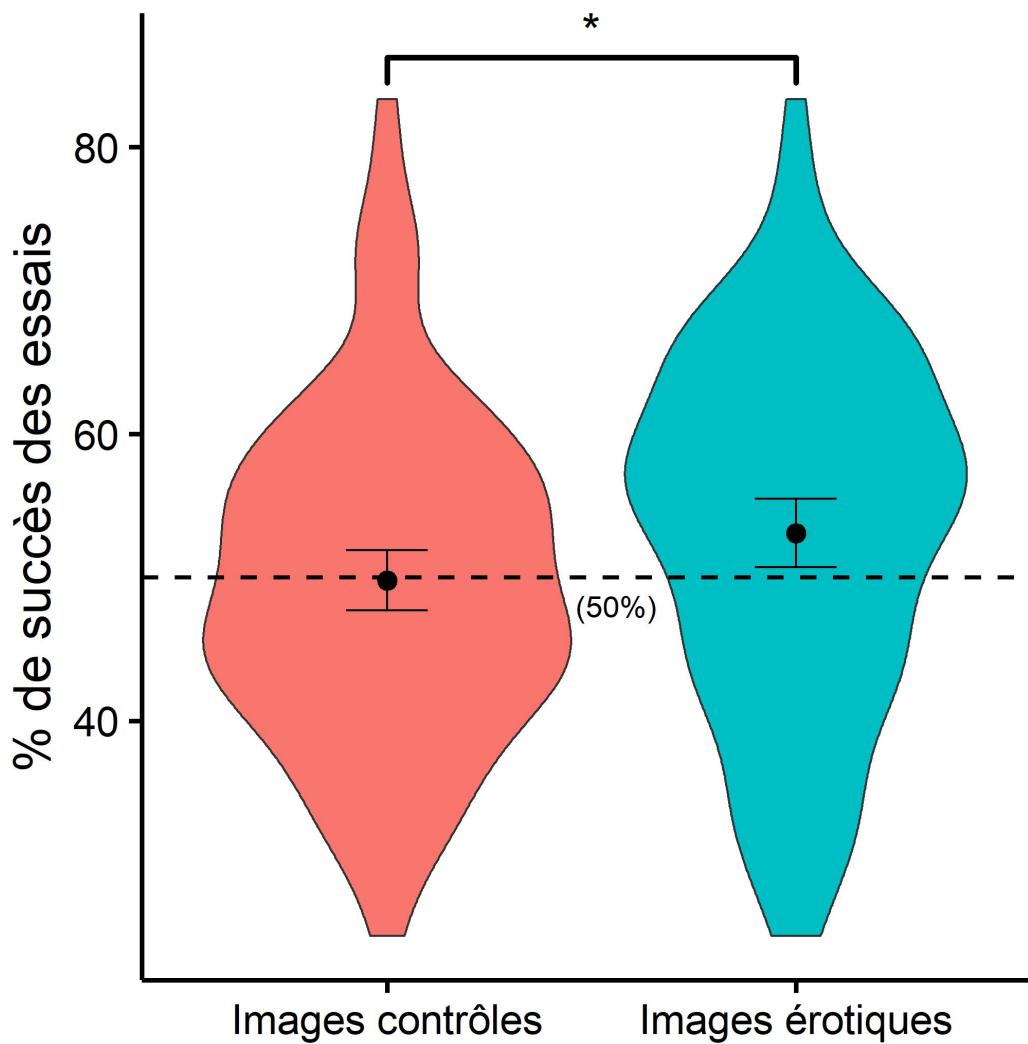
Tests t à échantillon unique : compare la moyenne d'un groupe à une valeur fixe connue (ici : le taux de chance de base de 50%).

On compare ce score pour les essais où une photo érotique apparaît :

$M = 53.1\%$ (vs 50%), $t(99) = 2.51$, $p = .01$, $d = 0.25$.

Et le score pour les essais où une photo non-érotique apparaît :

~~$M = 49.8\%$ (vs 50%), $t(99) = 0.15$, $p = .56$~~



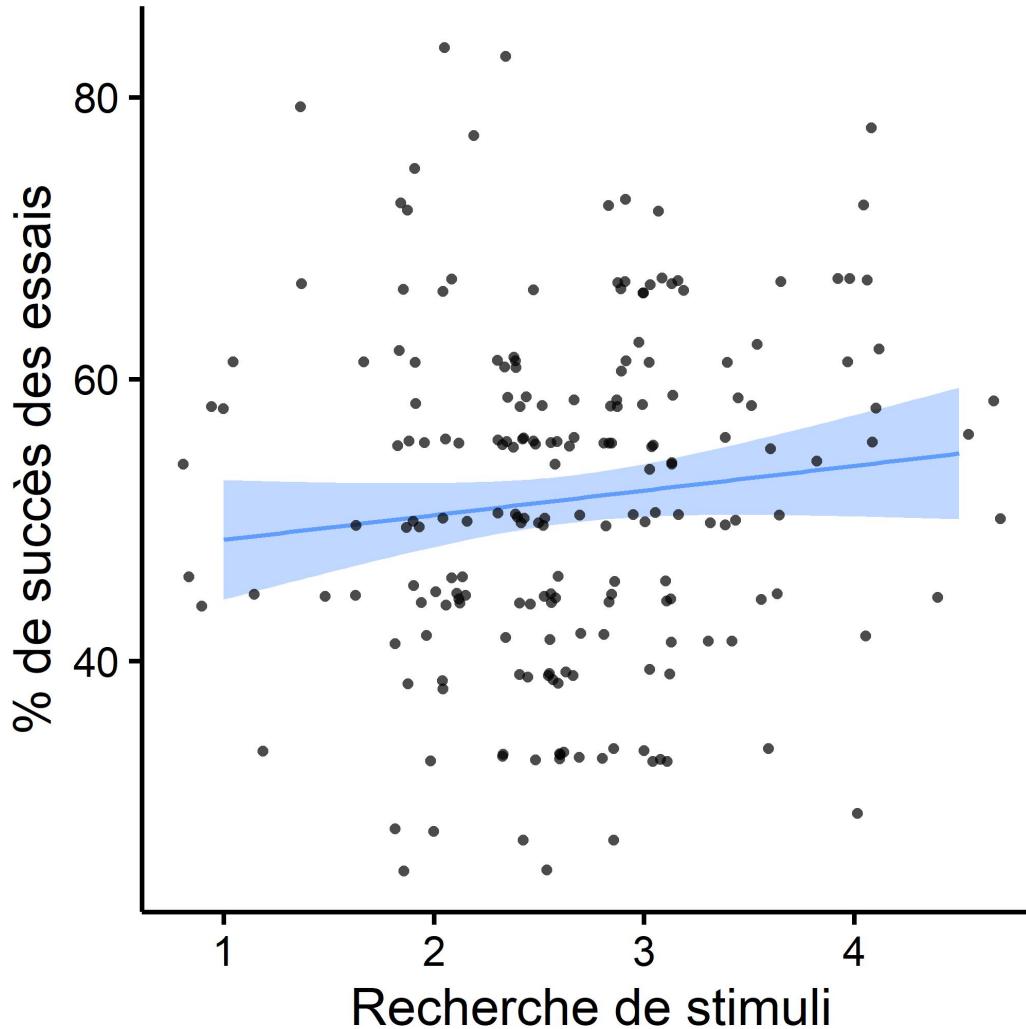
1) Détection précognitive de stimuli érotiques

Test *t* pour échantillons appariés : compare la moyenne des mêmes participants dans deux conditions (ici : photos érotiques vs non-érotiques).

$$t_{\text{diff}}(99) = 1.85, p = .031, d = 0.19$$

Corrélation : relation entre deux variables (ici : performance psi et le trait de recherche de *stimuli*).

$$r = .18, p = .035$$



2) Évitement précognitif des stimuli négatifs

- 36 essais/participant.es tirées du IAPS
- Présentation d'image neutres en miroirs, puis exposition à des images positives ou négatives 33 ms
- Succès = Préférer la cible qui permet d'éviter l'image négative.
- **Hypothèse : Les participant.es vont significativement plus que la chance éviter les images à valence négative.**
- **Mesure = Nombre d'image non- associée à l'exposition négative évitée**



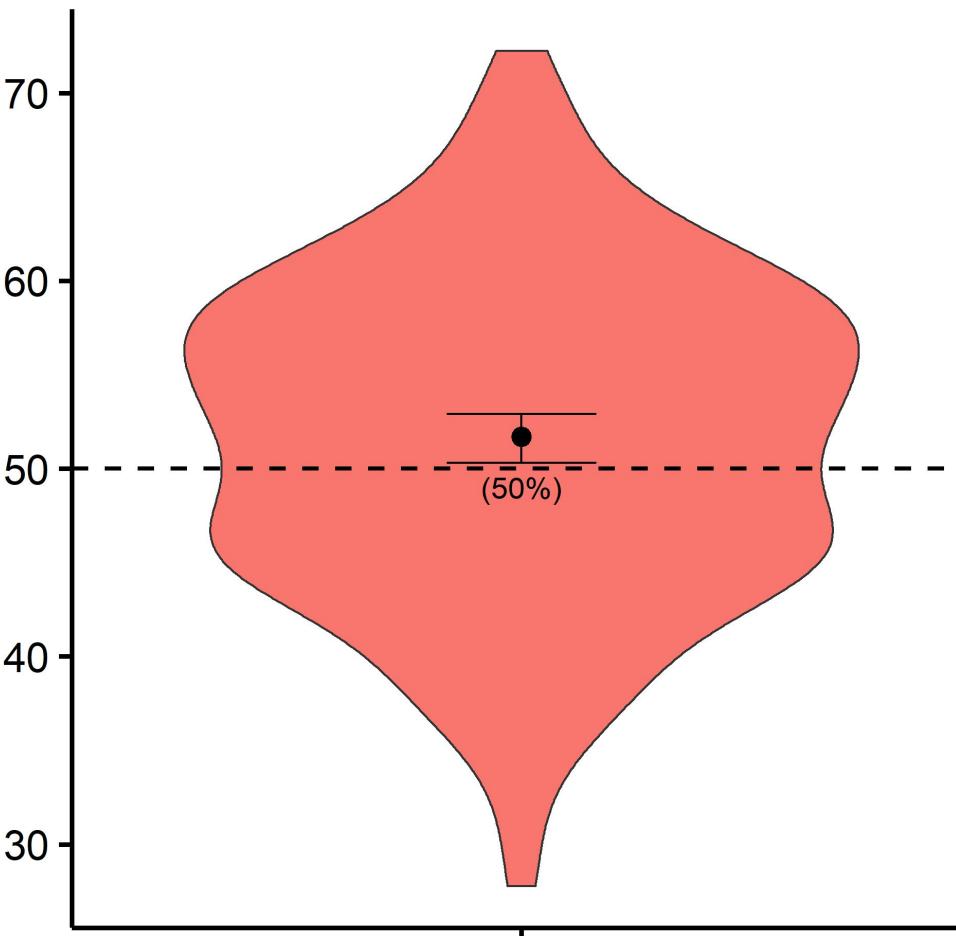
2) Évitement précognitif des stimuli négatifs

Hypothèse : Les participants préféreront la cible à la non-cible dans significativement plus de 50% des essais car les gens sont motivés à éviter les stimuli négatifs explicites.

Test t à échantillon unique : compare la moyenne d'un groupe à une valeur fixe connue (ici : le taux de chance de base de 50%).

$M = 51.7\%$ (vs 50%), $t(149) = 2.39$, $p = .009$, $d = 0.20$

% de succès des essais



Évitement précognitif des stimuli négatifs

2) Évitement précognitif des stimuli négatifs

Corrélation : relation entre deux variables (ici : performance psi et le trait de recherche de stimuli).

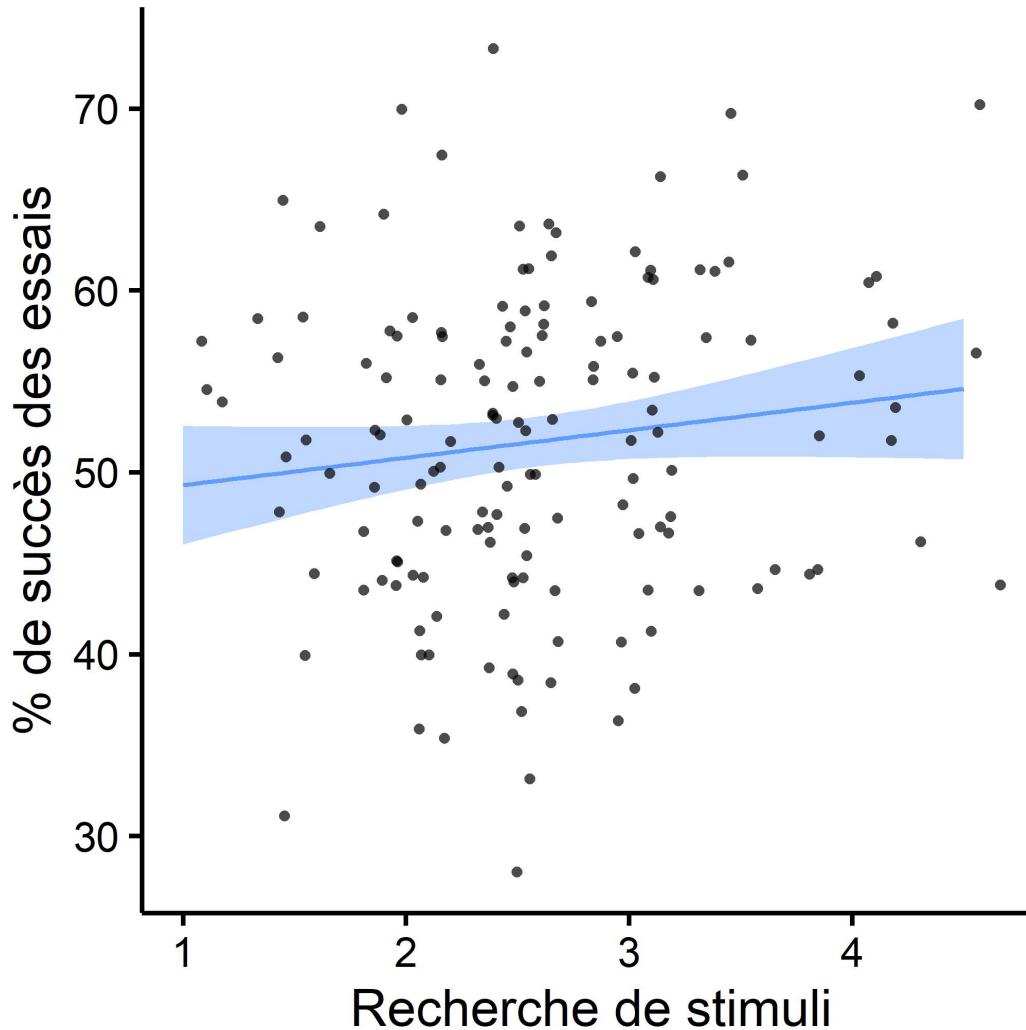
$$r = .17, p = .02$$

Et le score pour les participants hauts en recherche de stimuli :

$$M = 53.5\% \text{ (vs } 50\%), t(49) = 3.07, p = .002, d = 0.43$$

Et le score pour les participants bas en recherche de stimuli :

~~$$M = 50.7\% \text{ (vs } 50\%), t(99) < 1, p > .18$$~~



3) Amorçage rétroactif I

- Amorçage avant que l'image apparaisse
- Participant.es jugent aussi vite que possible si une image est plaisante ou pas plaisante
- **Mesure = Temps de réponse**
- **Hypothèse = Amorçage en congruence avec image = Plus grande rapidité d'exécution**
- Inversement de la séquence temporelle, le prime arrive après plutôt qu'avant que les jugements furent faits
- 32 essais avec amorçage rétroactif, puis 32 essais avec amorçage standard

A close-up photograph of a pink lotus flower in bloom, with its petals transitioning from light pink at the base to a deeper shade towards the edges. The flower is supported by a large, vibrant green leaf in the foreground. The background is blurred, showing what appears to be a traditional tiled roof.

Bonheur

3) Amorçage rétroactif I

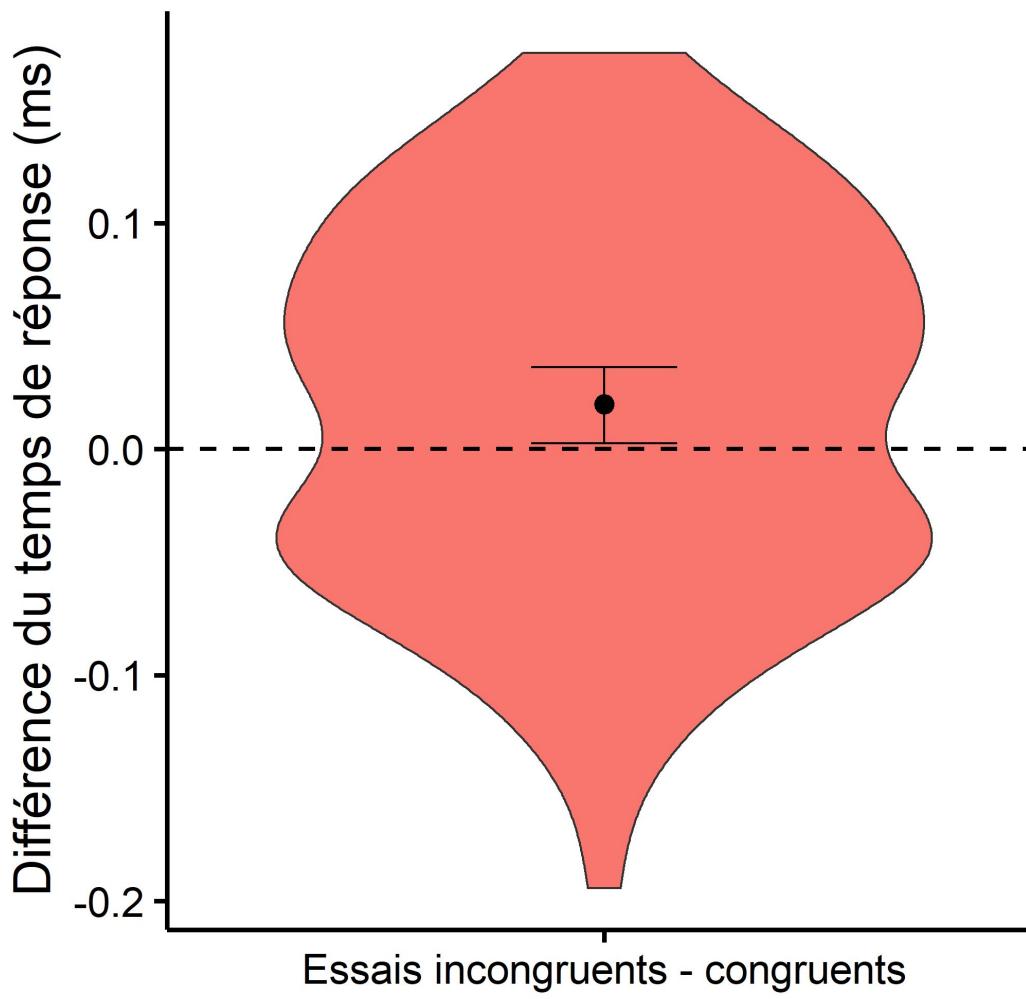
Hypothèse : Les participants répondront plus vite aux essais congruents qu'aux essais incongruents car il y aura moins de conflit cognitif.

Test t à échantillon unique : compare la moyenne d'un groupe à une valeur fixe connue (ici : différence à priori de 0 ms entre essais congruents et incongruents).

M = 15.0 ms (plus vite), t(96) = 2.55, p = .006, d = 0.25

Corrélation : relation entre deux variables (ici : performance psi et le trait de recherche de *stimuli*).

~~r = -.05, p > .05~~



4) Amorçage rétroactif II

- Expérience similaire que l'amorçage rétroactif I
- Avant l'expérience, pour chaque image était associée un mot à valence positif et un mot à valence négatif.
- Ce sont ces mots qui étaient choisis au hasard, plutôt que les images.

4) Amorçage rétroactif II

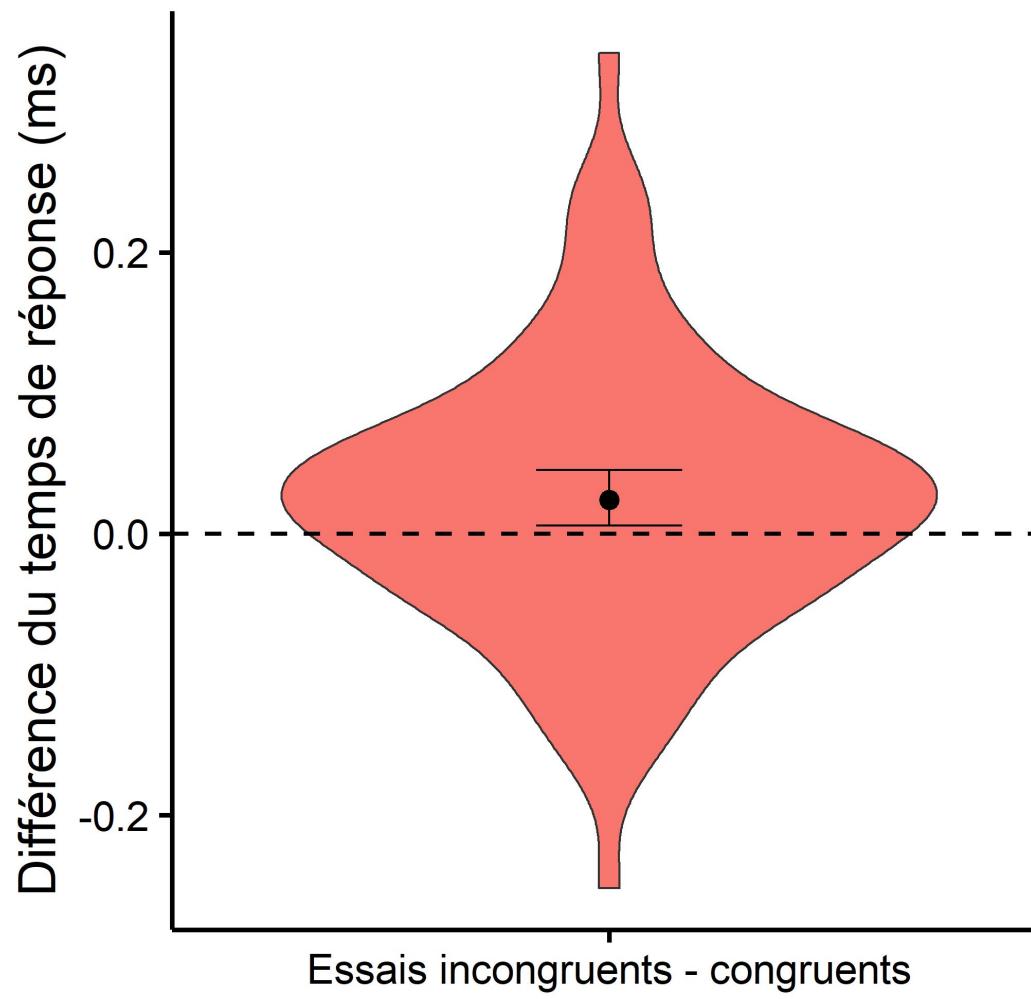
Hypothèse : Les participants répondront plus vite aux essais congruents qu'aux essais incongruents car il y aura moins de conflit cognitif.

Test t à échantillon unique : compare la moyenne d'un groupe à une valeur fixe connue (ici : différence à priori de 0 entre essais incongruents et congruents).

M = 16.5 ms (plus vite), t(98) = 2.03, p = .023, d = 0.20

Corrélation : relation entre deux variables (ici : performance psi et le trait de recherche de *stimuli*).

~~r = -.07, p > .05~~



5-6) Habituation rétroactive I et II

- Le participant est exposé des images subliminales, nommée les cibles d'habituation. Le fait d'être habitué à une image fait perdre sa valeur émotionnelle (arousal).
- Le participant voit deux images côté à côté, et indique laquelle il préfère. Une des images est la cible d'habituation. Si le participant préfère la cible, c'est un succès
- **Mesure = Nombre d'image préféré qui était la cible d'habituation**
- **Hypothèse = Le nombre de succès sera significativement plus élevé que la chance, et ce pour les images à valence négative (les images érotiques seront sous le seuil de la chance)**
- Dans l'habituation rétroactive II, un set d'image différent était utilisé pour les femmes et les hommes



5) Habituation rétroactive I

Hypothèse : Dans les essais avec des paires d'images négatives, les participants préféreront la cible à la non-cible dans plus de 50% des essais car l'habituation l'aura rendu moins négative.

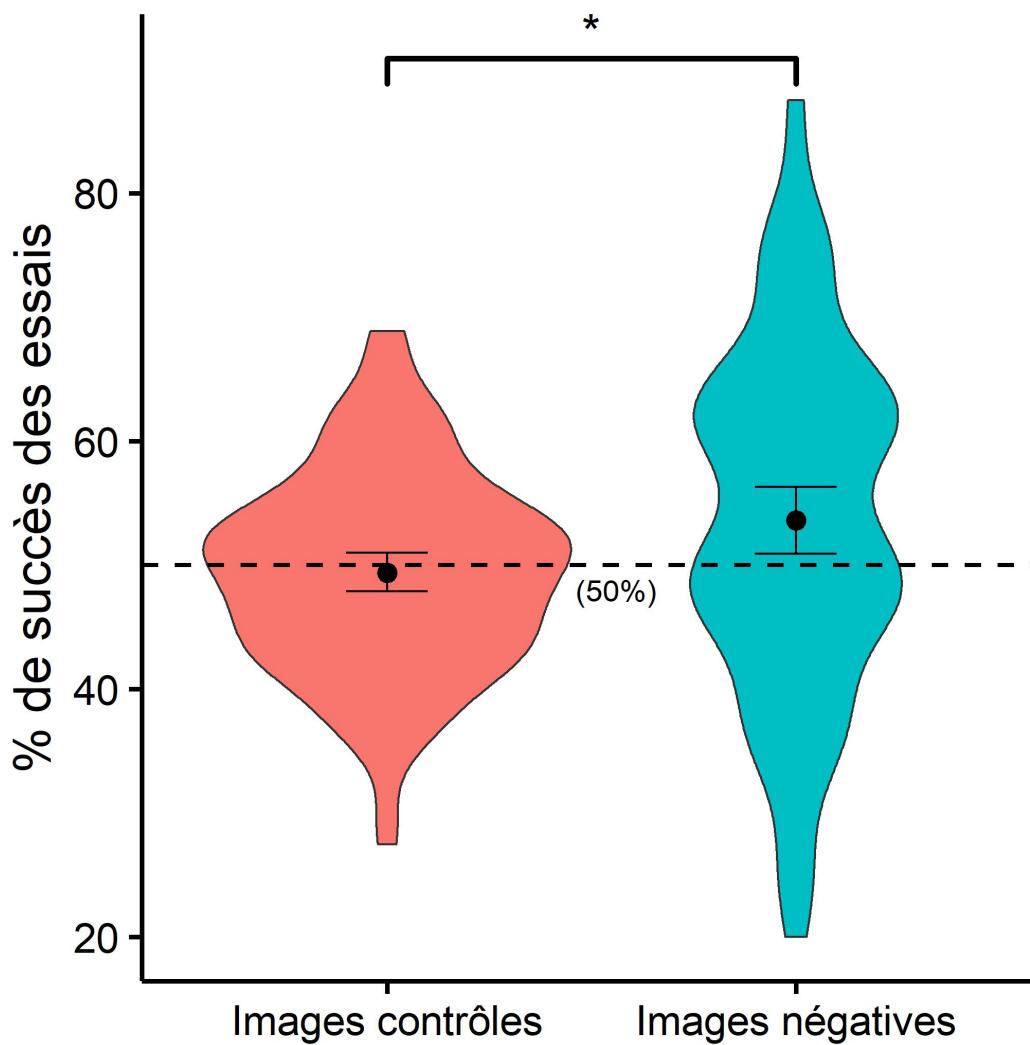
Tests t à échantillon unique : compare la moyenne d'un groupe à une valeur fixe connue (ici : le taux de chance de base de 50%).

On compare ce score pour les essais où des photos négatives apparaissent :

M = 53.1% (vs 50%), t(99) = 2.23, p = .014, d = 0.22

Et le score pour les essais où des photos neutres apparaissent :

~~M = 49.4% (vs 50%), t(99) = -0.73, p = .46~~



6) Habituation rétroactive II

Hypothèse : Les participants préféreront la photo cible (celle plusieurs fois exposée) dans plus de 50% pour les essais négatifs et moins de 50% des essais érotiques car leur valence émotionnelle sera réduite.

Tests t à échantillon unique : compare la moyenne d'un groupe à une valeur fixe connue (ici : le taux de chance de base de 50%).

On compare ce score pour les essais où des photos négatives apparaissent :

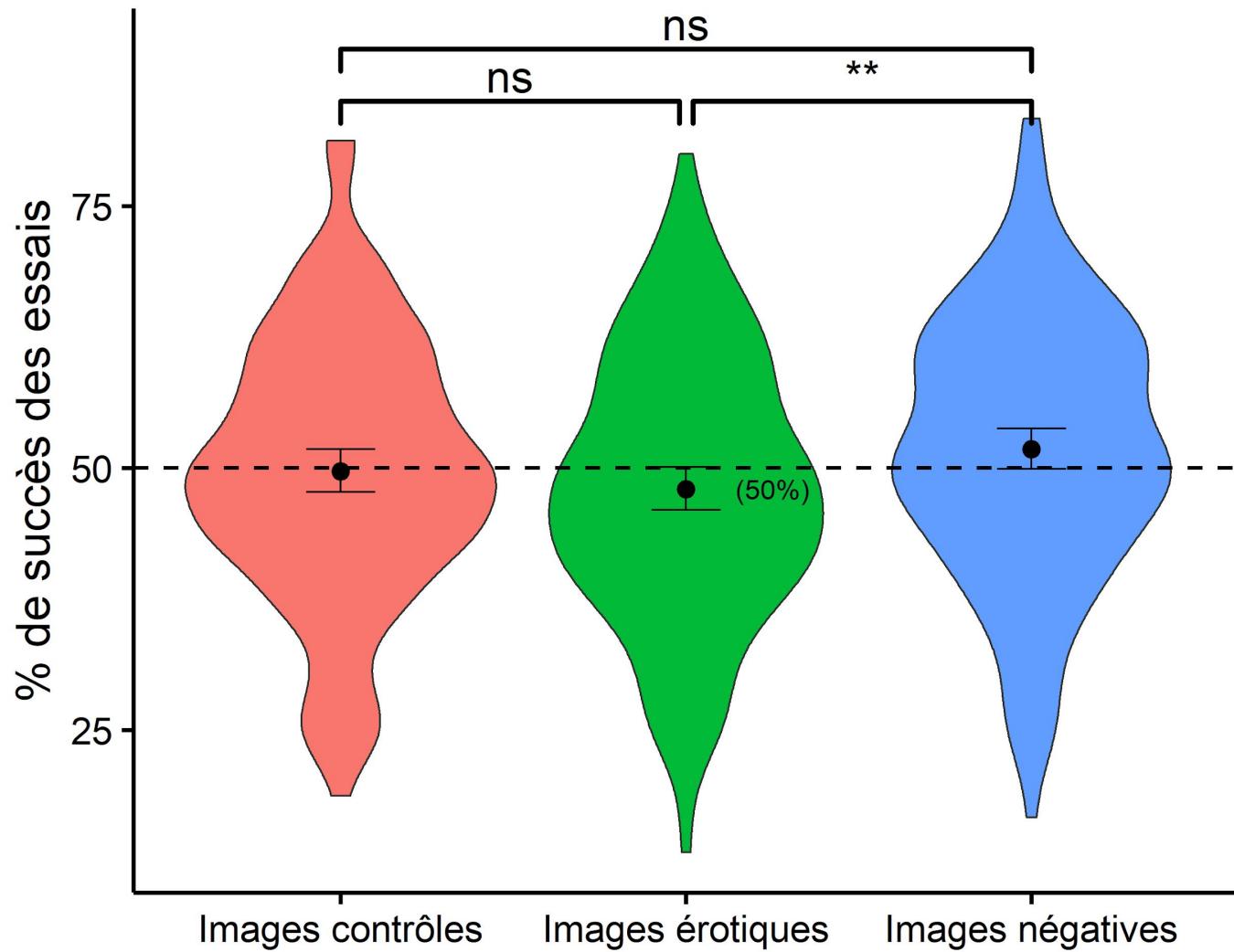
M = 51.8% (vs 50%), t(149) = 1.80, p = .037, d = 0.15

Et le score pour les essais où des photos érotiques apparaissent :

M = 48.2% (vs 50%), t(149) = 1.77, p = .039, d = 0.14

Et le score pour les essais où des photos neutres apparaissent :

~~**M = 49.3% (vs 50%), t(149) = 0.66, p = .51**~~



6) Habituation rétroactive II

Corrélation : relation entre deux variables (ici : performance psi et le trait de recherche de *stimuli*).

$$r = .24, p = .008$$

Et le score pour les participants hauts en *recherche de stimuli* :

$$M = 43.1\% \text{ (vs } 50\%), t(31) = 3.20, p = .002, d = 0.57$$

Et le score pour les participants bas en *recherche de stimuli* :

~~$$M = 51.3\% \text{ (vs } 50\%), t(67) = 0.78, p = .78, d = 0.09$$~~

7) Induction rétroactive de l'ennui

- Après réflexion sur l'expérience 6.
- Similaire à l'habituation rétroactive, mais l'image est présentée jusqu'à ennui.
- Présentation de manière subliminale d'images jusqu'à ennui.
- **Mesure = Nombre d'image préférée qui n'ont pas été présentée de manière subliminale**
- **Hypothèse = Le nombre d'image préférée sera significativement plus grand que la chance**

7) Induction rétroactive d'ennui

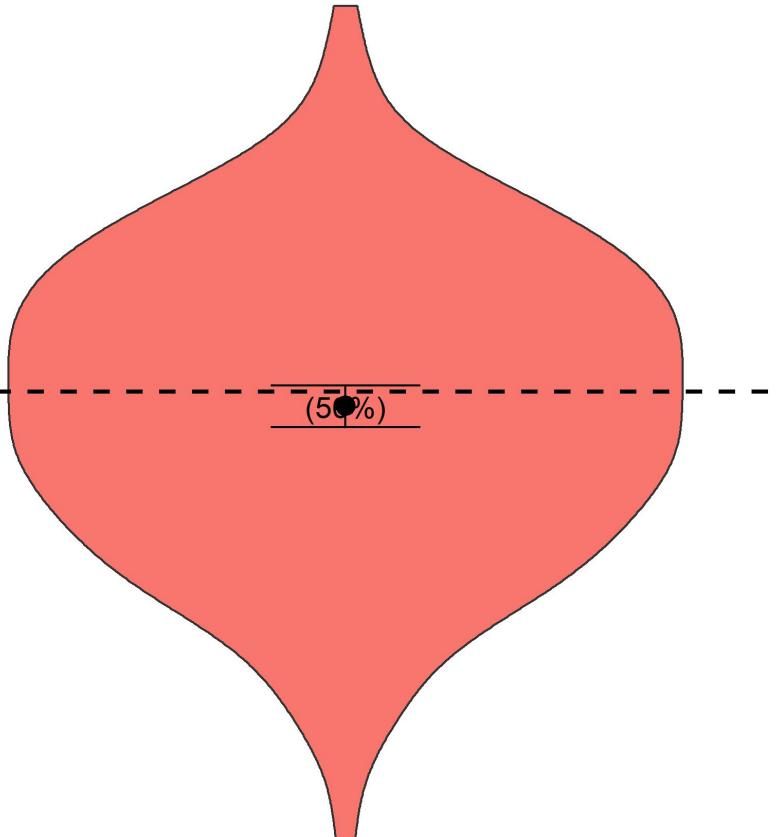
Hypothèse : Les participants préféreront la photo cible (celle maintes fois répétée) dans moins de 50% des essais car celle-ci sera devenue « ennuyeuse ».

Test t à échantillon unique : compare la moyenne d'un groupe à une valeur fixe connue (ici : le taux de chance de base de 50%).

~~M = 49.1% (vs 50%), t(199) = -1.31, p = .096, d = 0.09~~

% de succès des essais

60
40
20



Induction rétroactive d'ennui

7) Induction rétroactive d'ennui

Corrélation : relation entre deux variables (ici : performance psi et le trait de recherche de *stimuli*).

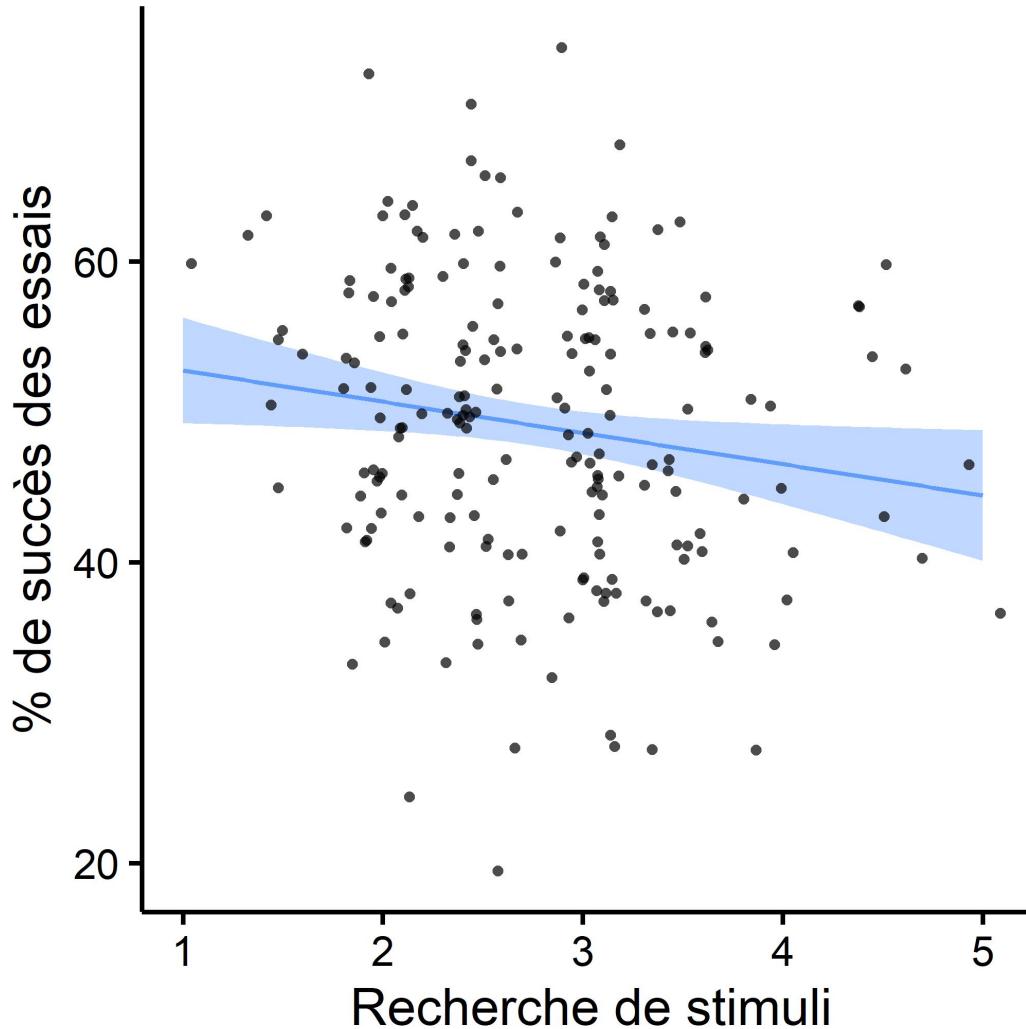
$$r = .16, p = .01$$

Et le score pour les participants hauts en *recherche de stimuli* :

$$M = 47.9\% \text{ (vs } 50\%), t(95) = 2.11, p = .019, d = 0.22$$

Et le score pour les participants bas en *recherche de stimuli* :

~~$$M = 50.1\% \text{ (vs } 50\%), t(103) = 0.17, p = .43$$~~



8-9) Facilitation du rappel rétroactif I et II

- Les participant.es recevaient un set de mots, devaient faire une tâche de rappel.
- Puis, on leur demandait de se pratiquer à se remémorer les mots
- 48 mots de quatre catégories (24 pratiqués et 24 non-pratiqués)
- **Mesure = Nombre de mots retenus**
- **Hypothèse = Le nombre de mots retenu sera significativement plus élevé pour ceux qui ont été pratiqué après le test.**

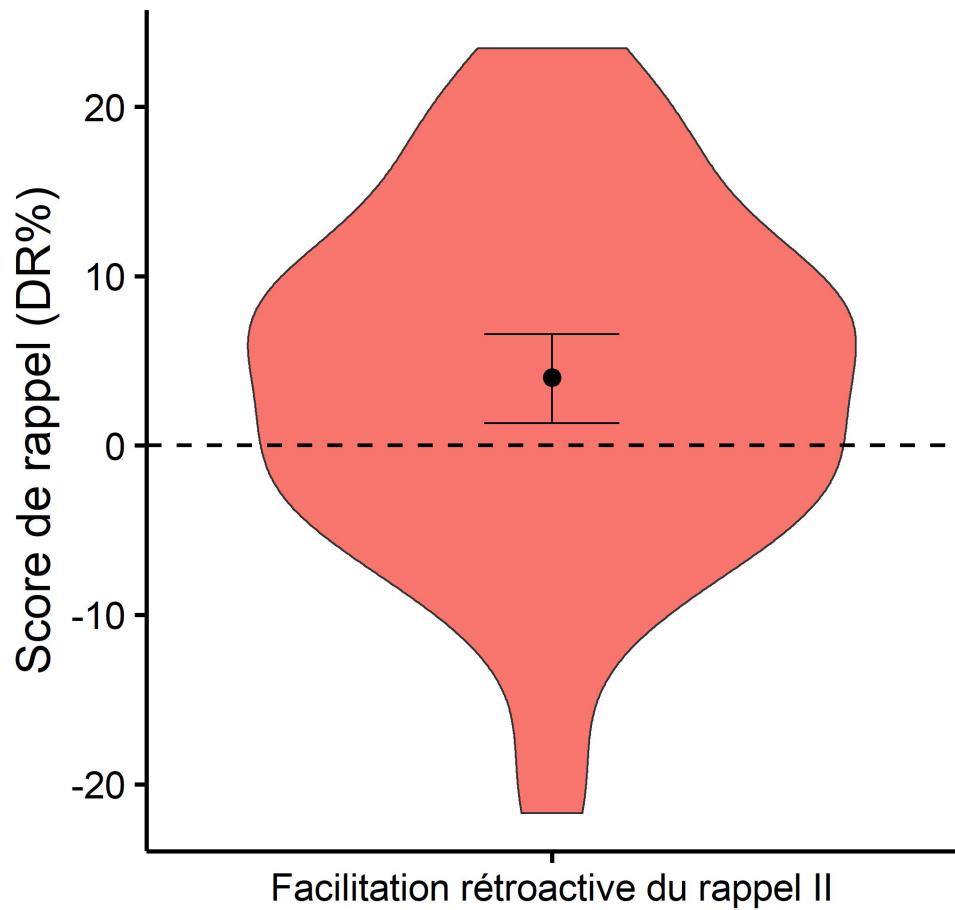
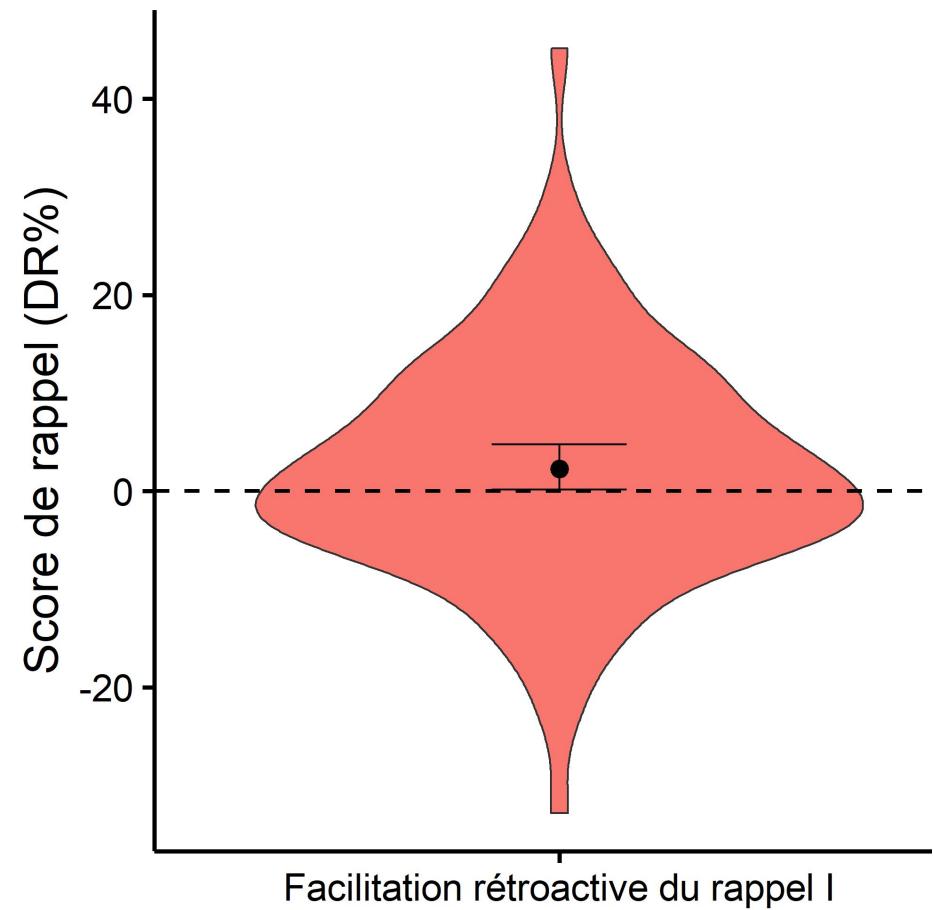
8-9) Facilitation rétroactive du rappel I

Hypothèse : Les participants se rappeleront plus des mots-à-être-pratiqués que les mots qui ne seront pas pratiqués.

Test t à échantillon unique : compare la moyenne d'un groupe à une valeur fixe connue (ici : le taux de chance de base de 0% de rappel).

8 : $M = 2.27\%$ (vs 0%), $t(99) = 1.92$, $p = .029$, $d = 0.19$

9 : $M = 4.21\%$ (vs 0%), $t(49) = 2.96$, $p = .002$, $d = 0.42$



8) Facilitation rétroactive du rappel I

Corrélation : relation entre deux variables (ici : performance psi et le trait de recherche de stimuli).

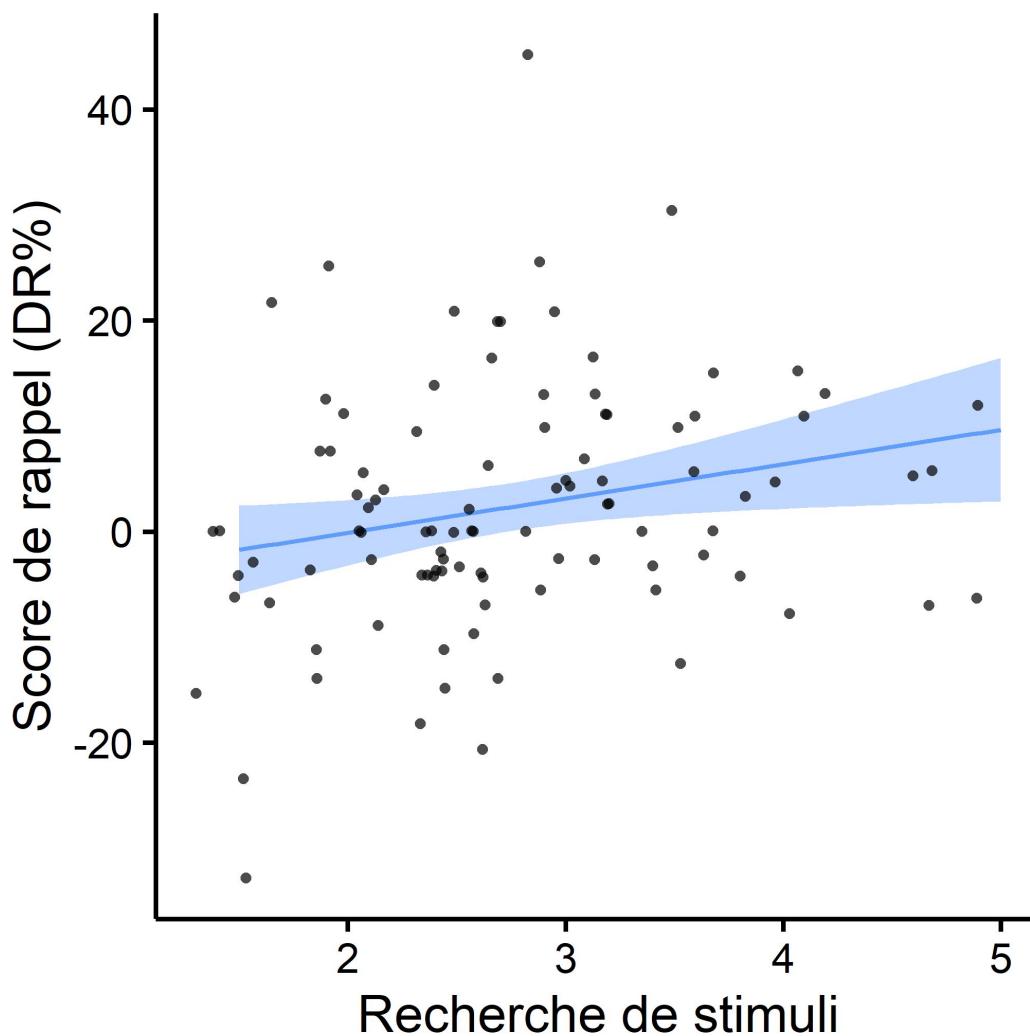
$$r = .22, p = .014$$

Et le score pour les participants hauts en recherche de stimuli :

$$M = 6.46\% \text{ (vs } 0\%), t(42) = 3.76, p = .0003, d = 0.57$$

Et le score pour les participants bas en recherche de stimuli :

~~$$M = 0.90\% \text{ (vs } 0\%), t(56) = 0.60, p = .73$$~~



9) Facilitation rétroactive du rappel II

Corrélation : relation entre deux variables (ici : performance psi et le trait de *recherche de stimuli*).

$$\cancel{r = .10, p = .25}$$

Le score pour les participants hauts en *recherche de stimuli* :

$$M = 4.47\% \text{ (vs } 0\%), t(15) = 1.77, p = .049, d = 0.49$$

Le score pour les participants bas en *recherche de stimuli* :

$$M = 4.09\% \text{ (vs } 0\%), t(33) = 2.34, p = .013, d = 0.40$$

Résumé des résultats

Table 7

Psi Performance in All Nine Experiments: Probability Levels (p), Effect Sizes (d), and Correlations (r) With Stimulus Seeking (SS)

Phenomenon tested and experiment	p full sample	d full sample	Correlation with SS	p high SS	d high SS	p low SS	d low SS
Precognitive approach/avoidance							
1. Detection of Erotic Stimuli	.01	0.25	.18*	.00002	0.71	.524	-0.01
2. Avoidance of Negative Stimuli ^a	.009	0.20	.17**	.001	0.45	.215	0.08
Retroactive priming							
3. Retroactive Priming I ^a	.007	0.26	-.05	.148	0.17	.036	0.24
4. Retroactive Priming II ^a	.014	0.23	-.07	.059	0.27	.035	0.23
Retroactive habituation							
5. Retroactive Habituation I Negative trials ^b	.014	0.22					
6. Retroactive Habituation II Negative trials ^b	.037	0.15					
Erotic trials	.039	0.14	0.24***	.002	0.57	.219	-0.09
7. Retroactive Induction of Boredom ^a	.096	0.09	.16**	.018	0.22	.483	0.00
Retroactive facilitation of recall							
8. Facilitation of Recall I	.029	0.19	.22**	.0003	0.57	.525	-0.08
9. Facilitation of Recall II	.002	0.42	-.10	.049	0.41	.013	0.40
Mean effect size (d)		0.22			0.43		0.10

^a Probabilities and effect sizes in this row are based on the mean of the *t* values across the variations of the data analysis. ^b The Stimulus-Seeking Scale was not administered in this experiment.

* *p* < .05. ** *p* < .02. *** *p* < .01.

Discussion

Est-ce vraiment de la précognition?

1. La rigueur en randomisation : les tests de L'écuyer (2001), Marsaglia (1995) et la battery de test of randomness de Diehard
2. Quatre hypothèses :
 - « Clairvoyance/remote viewing: The participant is accessing already determined information in real time, information that is stored in the computer.
 - » « Precognition or retroactive influence: The participant is, in fact, accessing information yet to be determined in the future, implying that the direction of the causal arrow has been reversed. »
 - « Psychokinesis: The participant is actually influencing the RNG's placements of the targets. »

Discussion (Félix)

« Artifactual correlation: The output from the RNG is inadequately randomized, containing patterns that fortuitously match participants' response biases. This produces a spurious correlation between the participant's guesses and the computer's placements of the target picture. »

Peut-on réellement dire que la causalité a été inversée?

CAUSE ← → EFFET

Diverses variables modératrices

- « Erotic stimuli seeking ». Les participant.es sont-ils tous sensibles ou à la recherche de l'érotisme?
- Zuckerman's sensation seeking scale (1974) utilisé dans l'habituation rétroactive
- Deux expériences de réplication : Savva, Child, and Smith (2004) et Parker et Sjoden (2010)
- Utilisation de contrôle groupe pour l'expérience de rappel (8 et 9) : quel est le rôle de la catégorisation (clustering?)? ET n'était-ce que de la clairvoyance?

Discussion

- Tester le PSI dans des environnements bien contrôlés
- Rendre possible et encourager la réPLICATION
- Utiliser des tests simples, connus et accessibles
- « Simplicity and familiarity become essential tools of persuasion. »
- Utilisation de variations pour écarter certaines hypothèses
- « Like most social-psychological experiments, the experiments reported here required extensive pilot testing. As all research psychologists know, many procedures are tried and discarded during this process. This raises the question of how much of this pilot exploration should be reported to avoid the file-drawer problem, the selective suppression of negative or null results. »

Critique de la validité

Externe

- A. Rappel : Biais de sélection sur l'échantillonnage
- B. Validité faible → Tâches de laboratoire peu généralisable
- C. Validité « Outcome » → élevé, considérant 9 tâches différentes

Interne

- A. Hypothèses alternatives soulevés et rejetées
- B. Plusieurs Mesures quantitatives fortes et simples → attrition, fidélité inter-juge, consistance interne (alpha) et temporelle ne s'appliquent pas
- C. Excellentes conclusions statistiques
- D. Très parcimonieux, analyses et expériences simples et accessibles

Critique : Validité interne

Pour faire une inférence causale, il faut remplir trois conditions :

- 1) La cause et l'effet doivent covarier (✓)
- 2) Les hypothèses rivales doivent être invraisemblable (✓)
- 3) La cause doit précéder l'effet (✗)

Quoi d'autre peut avoir joué? (Félix)

- La maturation?
- L'effet de Hawthorne?
- Le biais de l'expérimentateur?
- L'événement historique ?

Critique : Validité de construit

Validité de construit très élevée car Bem réutilise les tâches cognitives les plus connues et utilisées en psychologie sociale et cognitive.

Cependant...

Validité conceptuelle

- Aucune bonne théorie ne permet de soutenir ou d'expliquer les effets observés
- La littérature n'est que peu convaincante
- Explications Évolutionnistes VS Quantique VS spirituelle VS ...
- L'appel à l'ignorance : « On n'a pas de preuve que ça existe pas! »
- L'explication de Bem : « On découvre des phénomènes, puis on les explique »

VD et critique

Variable dépendante :
performance psi

Mais difficile de prouver qu'il s'agit d'un effet psi... c'est par défaut « psi » si aucune autre explication n'est trouvée.



Effet d'expérimentateur... psi

L'expérience de Wiseman & Schlitz (1998) sur les effets d'expérimentateur et la détection à distance du regard

Schlitz = croyant au psi

Wiseman = sceptique



Critiques générales (1)

- C'est sûr qu'avec de grandes tailles d'échantillon, toutes petites différences ressortiront comme significatives...
- Devis intra-sujet réduit la variance et donne plus de puissance pour détecter de petits effets...
- Pas de corrections pour comparaisons multiples!
- Il a utilisé un test unidirectionnel alors qu'il aurait peut-être dû utiliser des tests bidirectionnels puisqu'il ne pouvait pas vraiment prédire la direction de l'effet (à moins qu'il est lui-même eu des capacités psychiques).

Critiques générales (2)

- Le trait de recherche de stimuli ne s'appliquait pas ou différemment selon les différentes études. N'est-ce que de la chance?
- Bem a peut-être sélectionné quels tests et variables rapporter après avoir fait une multitude de test, pour choisir principalement ceux qui étaient significatifs.
- Ex : Images érotiques et non-érotiques se divisent en quatre types : romantique, neutre, négatives, et érotiques... (trop de degré de liberté)

Forces de l'étude

- Aveugle double, car ni le participant ni l'expérimentateur ne savait dans quelle condition (répétée) le participant était à chaque moment.
- Tâches toutes automatisées, laissant peu de place à l'erreur ou l'influence de l'expérimentateur
- Devis et analyses statistiques simples (tests *t*)
- Devis expérimental fort protège contre la majorité des menaces à la validité
- Étude originale mais réutilisant les tâches les plus établies en psychologie
- Grande tailles d'échantillon
- Plusieurs études (9)

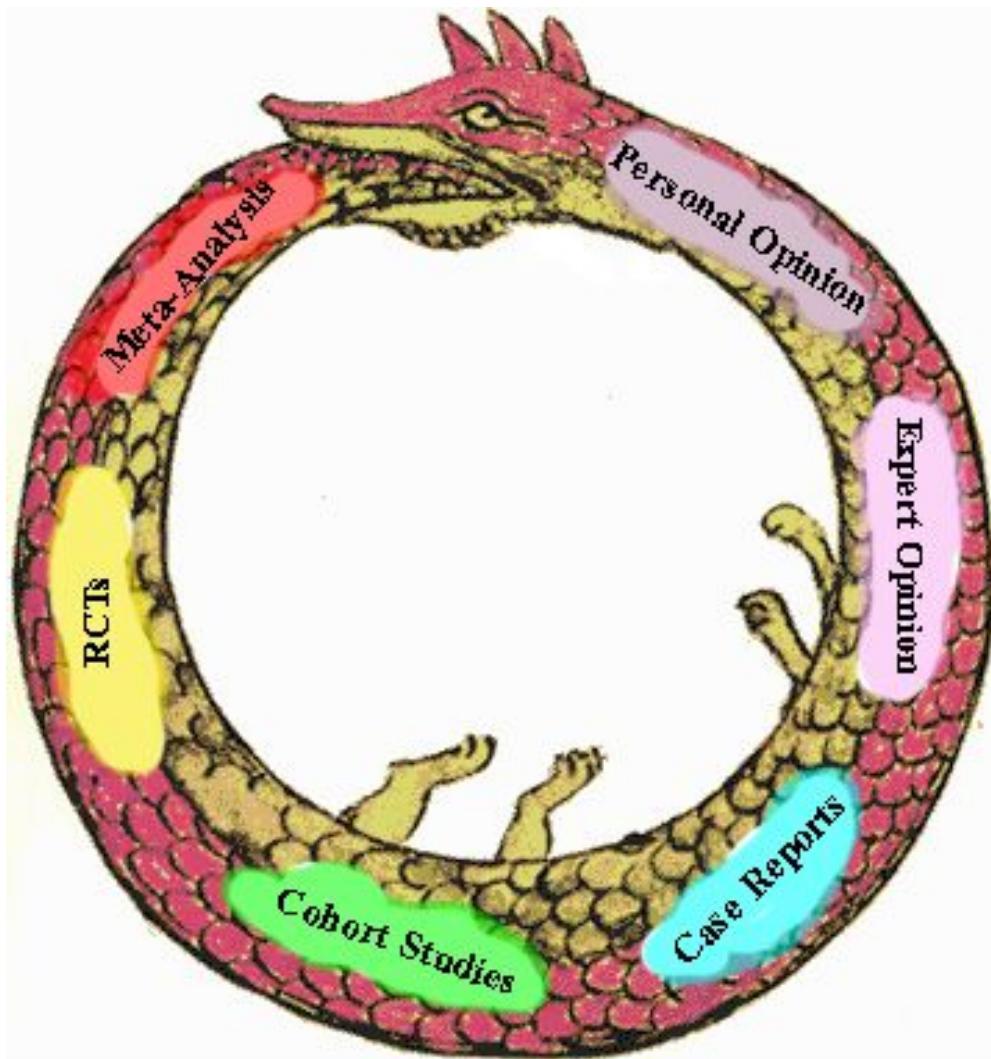


Check for updates

RESEARCH ARTICLE

REVISED Feeling the future: A meta-analysis of 90 experiments on the anomalous anticipation of random future eventsDaryl Bem¹, Patrizio E. Tressoldi ², Thomas Rabeyron^{3,4}, Michael Duggan⁵

We here report a meta-analysis of 90 experiments from 33 laboratories in 14 countries which yielded an overall effect greater than 6 sigma, $z = 6.40$, $p = 1.2 \times 10^{-10}$ with an effect size (Hedges' g) of 0.09. A Bayesian analysis yielded a Bayes Factor of 5.1×10^9 , greatly exceeding the criterion value of 100 for “decisive evidence” in support of the experimental hypothesis.



Pratiques de recherche questionnables

1. Plusieurs (trop de) variables dépendantes
2. Ne pas rapporter toutes les conditions
3. Arrondir de manière généreuse
4. P-Hacking et HARKing (« Hypothesizing After Results are Known »)
5. Exclusion de certaines données
6. Arrêter la collecte de données plus tôt que prévu (quand on a un effet)/snooping
7. Rapporter seulement les expériences significatives et déclarer les autres comme « pilotes » (*effet tiroir*)

Les bonnes pratiques en recherche I

- Avoir des très grandes tailles d'échantillon
- Avoir des valeurs p plus conservatrices
- Corriger pour les comparaisons multiples
- Toujours inclure les tailles d'effet et pas seulement les valeurs p
- Des critères plus strictes pour les tailles d'effet
- Faire des études de haute qualité comme des Essais Randomisés Contrôlés (ERC)
- Mettre ses données en ligne pour faciliter la réPLICATION statistique par des chercheurs indépendants et l'inclusion dans les méta-analyses
- Mettre tout son matériel expérimental accessible en ligne pour faciliter les réPLICATIONS indépendantes
- Avoir plusieurs réPLICATIONS (préférablement exactes)

Les bonnes pratiques en recherche II

- Se fier à des méta-analyses plutôt qu'à des études individuelles
 - Études homogènes (ou prendre en compte l'hétérogénéité)
 - Valeurs p normalement distribués (démontrer l'absence de biais)
- Utiliser des statistiques bayésiennes plutôt que fréquentistes (ou les deux)
- Adresser les biais de publication/effets tiroirs par des analyses statistiques poussées (e.g., « funnel plots »)
- Pré-enregistrer son étude en ligne (méthodologie, hypothèses, échantillon, critères d'exclusion, analyses statistiques) pour prévenir le « p -hacking »
- Sinon, clairement séparer les études et analyses **EXPLORATOIRES** (qui informeront les hypothèses de FUTURES études) versus les études et analyses **CONFIRMATOIRES** (qui visent à confirmer l'existence de phénomènes)

Réflexions épistémologiques

- Pourquoi cet article est si saisissant? Prendre un pas de recul.
- Doing science, writing science (Shickore, 2008)
- Paradigme scientifique (Kuhn, 1962)
- Différence entre statistiquement significatif et pratiquement significatif
- Principe de Falsification ou de réfutabilité (Popper, 1934)
- Le cul de sac des recherches sur le paranormal

Plan de recherche alternatif

- Méthodologie, hypothèses, échantillon, critères d'exclusion, analyses statistiques tous préenregistrés sur *Open Science Frameworks!!!*
- 2 phases
 - Phase 1 (entraînement des capacités psi)
 - Phase 2 (test des capacités psi)
 - Sujets doués pour trouver des effets plus forts et non-ambigus
 - Raisonnable d'assumer que les capacités psi sont normalement distribués dans la population
- Variable dépendante : performance psi

Nouvelle question de recherche

- Problème : vieille question de recherche N'EST PAS FALSIFIABLE
- Donc nouvelle question de recherche doit être falsifiable

Question de recherche additionnelle : Peut-on entraîner les capacités psi?

Plan de recherche alternatif : Méthode

Phase 1

- Entraînement des capacités psi
- Essai randomisé contrôlé (ERC) avec trois groupes:
 - **Groupe 1:** Liste d'attente sans intervention
 - **Groupe 2:** Pratique des tâches de Bem dans l'ordre temporel normal
 - **Groupe 3:** Pratique des tâches de Bem dans l'ordre temporel inverse

Participants phase 1

- Comme Bem : analyse de puissance
 - Taille d'effet de 0.25 (*d* de Cohen)
 - 80% de puissance
 - Donc 100 participants par groupe!
 - Ni plus, ni moins!! Pas 50, pas 150, ni 200!!
- Participants aléatoirement sélectionnés parmi toute la population de 18-35 ans sans condition psychiatrique ou neurologique

Groupe « Liste d'attente »

Phase 1

- **Temps 1** : Fait les mêmes tâches que Bem (mesure de base)
- **Intervention** : Pas d'intervention pendant 6 semaines
- **Temps 2** : Fait les mêmes tâches que Bem (mesure post)
- Contrôle pour la mesure de base, T2-T1
- Si les sujets doués sont vraiment doués et que leur performance n'est pas dûe à un artéfact statistique ou méthodologique, ils devraient conserver une bonne performance dans **TOUTES** les tâches de Bem
- Sinon peut-être dû à un artéfact statistique ou méthodologique

5 tâches (contrebalancées)

- 1) Détection précognitive de stimuli érotiques
- 2) Évitement précognitif des stimuli négatifs
- 3) Amorçage rétroactif
- 4) Habituation rétroactive
- 5) Facilitation rétroactive du rappel

Groupe pratique—ordre temporel *normal*

Phase 1

- **Temps 1** : Fait les mêmes tâches que Bem (mesure de base)
- **Intervention** : Se pratique sur les tâches de Bem (avec ordre temporel normal) à raison d'une heure par semaine pendant 6 semaines
- **Temps 2** : Fait les mêmes tâches que Bem (mesure post)
- Contrôle pour la mesure de base, T2-T1

Groupe pratique—ordre temporel *inversé*

Phase 1

- **Temps 1** : Fait les mêmes tâches que Bem (mesure de base)
- **Intervention** : Se pratique sur les tâches de Bem (avec ordre temporel inversé) à raison d'une heure par semaine pendant 6 semaines
- **Temps 2** : Fait les mêmes tâches que Bem (mesure post)
- Contrôle pour la mesure de base, T2-T1

Test statistiques phase 1

- Temps 2 moins Temps 1 pour contrôler pour la valeur de base
 - On ne conserve que la différence ou « l'amélioration »
- **Régression multiples avec contrastes planifiés**
 - Groupe ordre inversé vs groupe liste d'attente
 - Groupe ordre normal vs groupe liste d'attente
 - Groupe ordre inversé vs groupe ordre normal

Pas d'anova (test omnibus) puisque nous ne sommes pas intéressés à voir s'il y a des différences globales mais bien si nos groupes diffèrent spécifiquement

Alpha de .05, correction pour comparaison multiples (Bonferroni, correction plus conservatrice)

Hypothèses statistiques phase 1

Hypothèse 1 : Le groupe avec ordre de pratique *inversé* aura un taux de succès psi significativement plus élevé que les deux autres groupes (pratique avec ordre normal et liste d'attente).

Hypothèse 2 : Les groupes de pratique avec ordre normal ou liste d'attente auront une performance psi similaire.

Plan de recherche alternatif : Méthode

Phase 2 (BONUS)

- RéPLICATION des expériences de Bem avec tous les sujets *doués* de la phase 1 (pas tous ceux ayant été entraînés!)
 - + groupes contrôles pour chaque expérience
 - Si impossible d'entraîner cette capacité ou pas assez de sujets doués inclus, sujets doués naturels identifiés dans la population (assumant que c'est normalement distribué)
- Comme Bem, une taille d'échantillon de 100 participants par groupe (2 groupes : expérimental et contrôle)

Test statistiques phase 2

- Même tests que Bem :
 - Test t à échantillon unique
 - Test t pour échantillons appariés
 - Alpha de .05
 - MAIS: corrections pour comparaison multiples
(Bonferroni, correction plus conservatrice)
- + Test t pour échantillons indépendants

Hypothèses statistiques phase 2

Hypothèse : Les sujets psi doués auront une performance psi supérieure à la chance ainsi qu'au groupe contrôle.

Limites

- Phase 1
 - En prenant des groupes contrôle, on ajoute du bruit inter-sujet et de la variance, réduisant la puissance
 - On doit modifier la question de recherche un peu
 - Il n'y a aucune garantie que s'entraîner intentionnellement à cette tâche est la meilleure façon de développer ses capacités psi...
 - Longues tâches donc ennui donc attrition...
- Phase 2
 - On reste avec le problème de l'irréfutabilité concernant l'existence du psi
 - Pas d'assignation aléatoire aux groupes puisqu'on prend les sujets doués de la phase 1

Forces

Devis expérimental fort contrôle pour :

- Validité externe : meilleure généralisation
- Validité interne : parcimonie, analyse statistique, etc.
- Pratiquement significatif
- Ajout d'un groupe contrôle
- Évaluation inter et intra-sujet

Phase 2 permet une réPLICATION des performances des sujets doués de la phase 1

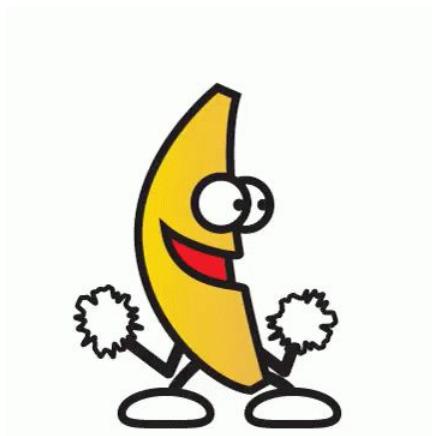
Conclusions

- Nous allons recruter les sujets doués dans cette classe pour de plus amples expérimentations ;)
- Repensez-y deux fois quand quelque chose d'apparence paranormale vous arrive
- Soyez attentif aux événements qui sortent de l'ordinaire

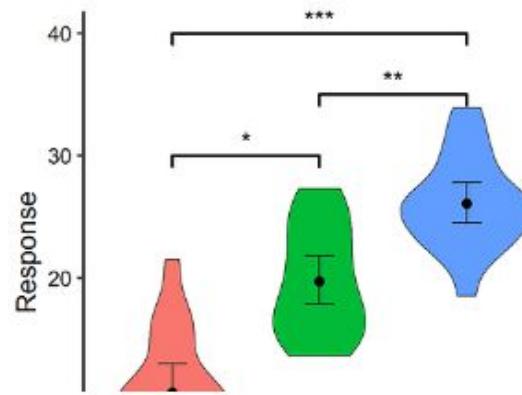
Mais surtout : pensez à avoir de bonnes pratiques de recherche!!!



Shameless Plug



<https://remitheriault.wixsite.com/site/blog>



NICE VIOLIN PLOTS

How to make nice-looking violin plots in R.

NICE SCATTER PLOTS

How to make nice-looking scatter plots in R.

