How subjective idea valuation energizes and guides creative idea generation

Alizée Lopez-Persem[1], Sarah Moreno Rodriguez[1], Marcela Ovando-Tellez[1], Théophile Bieth[1,2], Stella Guiet[1], Jules Brochard[3], Emmanuelle Volle[1]

1. FrontLab, Sorbonne University, Institut du Cerveau - Paris Brain Institute - ICM, Inserm, CNRS, AP-HP, Hôpital de la Pitié Salpêtrière, Paris, France.
2. Neurology department, Hôpital de la Pitié Salpêtrière, AP-HP, F-75013, Paris, France
3. TRA, "Life and Health", University of Bonn, Bonn, Germany

Corresponding authors: Alizée Lopez-Persem (lopez.alizee@gmail.com) & Emmanuelle Volle (emmavolle@gmail.com)

Author Contributions: EV, ALP designed the study. SMR, SG collected the data and performed model-free behavioral analyses. TB, MOT, SMR analysed the creativity battery data. ALP, EV, JB conceptualized the computational model. ALP performed the model-free and computational modelling analyses. ALP, EV wrote the article. All authors reviewed and edited the article.

Competing Interest Statement: Authors declare no competing interests.

Classification: Biological sciences

Keywords: Creativity, preferences, computational modelling, subjective value, idea generation, evaluation

This PDF file includes:

Main Text
Figures 1 to 7

Abstract

What drives us to search for creative ideas and why does it feel good to find one? While previous studies demonstrated the positive influence of intrinsic motivation on creative abilities, how reward and subjective values play a role in creative mechanisms remains unknown. The existing framework for creativity investigation distinguishes generation and evaluation phases, and mostly aligns evaluation to cognitive control processes, without clarifying the mechanisms involved. This study proposes a new framework for creativity by 1) characterizing the role of individual preferences (how people value ideas) in creative ideation and 2) providing a computational model that implements three types of operations required for creative idea generation: knowledge exploration, candidate ideas valuation (attributing subjective values), and response selection. The findings first provide behavioral evidence demonstrating the involvement of valuation processes during idea generation: preferred ideas are provided faster. Second, valuation depends on the adequacy and originality of ideas and determines which ideas are selected. Finally, the proposed computational model correctly predicts the speed and quality of human creative responses, as well as interindividual differences in creative abilities. Altogether, this unprecedented model introduces the mechanistic role of valuation in creativity. It paves the way for a neurocomputational account of creativity mechanisms.

Significance statement

How creative ideas are generated remains poorly understood. Here, we introduce the role of subjective values (how much one likes an idea) in creative idea generation and explore it using behavioral experiments and computational modelling. We demonstrate that subjective values play a role in idea generation processes, and show how these values depend on idea adequacy and originality (two key creativity criteria). Next, we develop and validate behaviorally a computational model. The model first mimics semantic knowledge exploration, then assigns a subjective value to each idea explored, and finally selects a response according to its value. Our study provides a mechanistic model of creative processes which offers new perspectives for neuroimaging studies, creativity assessment, profiling, and targets for training programs.

Main Text

1. Introduction

Creativity is a core component of our ability to promote change and cope with it. Creativity is defined as the ability to produce an object (or an idea) that is both original and adequate (1–3). Originality is critical to the concept of creativity; it refers to something novel or unprecedented. However, to be considered creative, a production also needs to be adequate. Adequacy corresponds to how appropriate, efficient to the goal a created entity is. The cognitive mechanisms underlying the production of an idea that is both original and adequate are yet to be elucidated. This study aims to decipher some of the cognitive processes of creative thinking by developing a new computational model, composed of three main operations: idea exploration, (e)valuation, and selection.

Creativity has been classically conceptualized and studied in neuroscience in the context of three main frameworks: the divergent thinking approach, the associative theory, and the insight problem-solving approach (4, 5). Generation tasks are typically used in those approaches, and the generated responses are overall assessed for originality and adequacy. It is largely admitted that creativity involves two interacting phases: generation and evaluation. Theoretical models including these two processes have been proposed (6, 7), such as the “two-fold model of creativity” (8), or the “blind-variation and selective retention model” (9, 10), a Darwinian-inspired theory stating that ideas are generated and evaluated on a trial and error basis, similarly to a variation-selection process. Additionally, neuroimaging findings support the distinction between generative and evaluative processes, notably with the involvement of the Default Network (DN) in relationship with generation and of the Executive Control Network (ECN) in relationship with evaluative and selection processes (11–13). However, what kind of processes underlies evaluation in the context of creativity (in other words, what evaluative processes drive selection) remains overlooked.

Previous frameworks assumed that the originality and adequacy of ideas are evaluated to drive the selection of an idea during idea production (14). Existing theories also usually align evaluation with controlled or metacognitive processes (i.e., monitoring and applying some control to select or inhibit early thoughts and adapt to the context) (6, 8, 11, 12, 15–19). However, how these processes work and result in idea selection remains unknown. Because evaluative processes in other domains involve subjective values that are assigned to options to guide selection (20), we hypothesize that evaluation in the context of creativity also requires building a subjective value. As previous work highlighted the importance of adequacy and originality in idea evaluation, we propose that this value is based on a combination of originality and adequacy of candidate ideas. Hence, we introduce valuation in the ideation process and dissociate them from other evaluation and generation processes. Valuation can be defined as a quantification of the subjective desire or preference for an entity (21) and consists in assigning a subjective value to an option, i.e., to define how much it is "likeable".

The neuroscience of value-based decision-making indeed demonstrated that valuation and other evaluation processes are distinct, experimentally dissociable, and have separate brain substrate (22, 23). Indeed, valuation processes have been investigated for centuries by philosophers, economists, psychologists, and more recently by neuroscientists (24), outside of the creativity field. Advances in the neuroscience of decision-making have allowed the identification of a neural network, the Brain Valuation System (BVS), representing the subjective value of options an agent considers (24). The BVS activity reflects values in a generic (independent of the kind of items) and automatic (even when we are engaged in another task) manner (25). Interestingly, the BVS is often coupled with the ECN when a choice has to be made: in a top-down manner – the ECN modulates values according to the context (26); and in a bottom-up way – by integrating decision-value, it drives the choice selection (27). The new framework that we propose through the present study is that the BVS is automatically involved during creativity, and that evaluation processes in creativity involve valuation, implemented by that network, in interaction with exploration and selection processes, supported by other networks.

Some studies have reported indirect arguments for the involvement of the BVS in creativity through a link with dopamine (28) or activation of the ventral striatum (16). Nevertheless, very little is known about the role of the BVS in creativity, and its interaction with the commonly reported brain networks (DN and ECN) for creativity has, to our knowledge, not been explored. In fact, the place of valuation processes in creativity still needs to be conceptualized and empirically investigated.

Here, we formulate the unprecedented hypothesis that originality and adequacy are combined into a "subjective value" according to individual preferences, and that this subjective value drives the creative degree of the output. This value can impact the selection of an idea, and also possibly have a motivational role (29) on the exploration of candidate ideas. Taking into account previous research from both creativity and decision-making fields, we hypothesize that creativity involves i) an *explorer* module that works on individual knowledge representations and provide a set of options/ideas varying in originality and adequacy; ii) a *valuator* module that computes the likeability of candidate ideas (their subjective value) based on a combination of their originality and adequacy with the goal an agent tries to reach; iii) a *selector* module that applies contextual constraints and integrates the subjective value of candidate ideas to guide the selection. To test these hypotheses, we combined several methods of cognitive and computational neuroscience. We build a computational model composed of the *explorer*, *valuator* and *selector* modules, that we modelled separately (Figure 1) as detailed below.

First, producing something new and appropriate (i.e. creative) relies in part on the ability to retrieve, manipulate or combine elements of knowledge stored in our memory (30, 31). Semantic memory network methods are a valuable approach to study these processes (32–35). Several semantic search mechanisms have been previously explored using for instance censored and biased random walks within semantic networks (36). The use of those models was essentially restricted to explaining fluency tasks (37) and retrieval of remote associates (38), but they have not yet been combined with decision models that could bring new insights into how individuals reach a creative solution. Based on this literature, we modelled the *explorer* module as a random walk wandering into semantic networks.

Second, valuation and selection processes are typically studied using decision models. Utility (economic term for subjective value) functions can well capture valuation of multi-attribute options that weigh attributes differently depending on individuals (39–41). Hence, we modelled the *valuator* module of our model as a utility function that assigns subjective values to candidate ideas based on the subjective evaluation of their adequacy and originality, considered as the necessary attributes of creative idea.

Third, the computed subjective value is then used to make a decision. Decision models assess value-based choices and can be static like softmax (42), dynamic like drift-diffusion models (43), or biologically inspired (44). Simple models like softmax functions can explain many types of choices, ranging from concrete food choices to abstract moral choices, as soon as they rely on subjective values. Here, we reasoned that such a simple function can capture and predict creative choices (*selector* module), when taking as an input subjective values of candidate ideas.

Overall, by means of different approaches to test our hypotheses, we developed an original computational model (Figure 1) in which each module (*explorer*, *valuator*, *selector*) was modelled separately. We aimed at 1) determining whether subjective valuation occurs during idea generation (creativity task) and defining a *valuator* module from behavioral measures during the decision-making tasks; 2) Developing the *explorer* and *selector* modules, and characterizing which module(s) relies on subjective valuation (*explorer* and/or *selector*); 3) Simulating surrogate data from the full model composed of the three modules and comparing it to human behavior; and 4) assessing the relevance of the model parameters for creative abilities.

2. Results

Sixty-nine subjects were included in the analyses (see Methods 4.1). The experiment consisted of several successive tasks (Figure 2, see Methods 4.2): the Free Generation of Associate Task (FGAT), designed to investigate generative processes and creative abilities, a likeability rating task, a choice task, an originality and adequacy rating task, and a battery of creativity assessment.

2.1. Subjective valuation in idea generation and development of the *valuator* module

2.1.1. FGAT behavior: the effect of task condition on speed and link with likeability

In the *First* condition of the FGAT task, participants were asked to provide the first word that came to mind in response to a cue. In the *Distant* condition, they had to provide an original, unusual, but associated response to the same cues as in the *First* condition (see Figure 2 and Methods 4.2.1).
We investigated the quality and speed of responses provided in the FGAT task in the *First* and *Distant* conditions. The quality of responses was investigated using their associative frequency obtained from the French database of word associations *Dictaverf* (see Methods 4.3.1), and using the ratings participants provided in three rating tasks requiring them to judge how much they liked an idea (likeability or satisfaction of a response to the FGAT *Distant* condition, see Methods 4.2.2), how much original they found it (originality), and how appropriate (adequacy).

FGAT responses: associative frequency
Consistent with the instructions of the FGAT conditions, we found that participants provided more frequent responses (i.e., more common responses to a given cue based on the French norms of word associations *Dictaverf*) in the *First* condition than in the *Distant* condition ($\log(\text{Frequency}_{\text{First}})=-3.25\pm0.11$, $\log(\text{Frequency}_{\text{Distant}})=-6.21\pm0.11$, M±SEM, $t(68)=18.93$, $p=8.10^{-29}$). Then, we observed that response time in the FGAT task decreased with the cue-response associative frequency, both in the *First* ($\beta=-0.34\pm0.02$, $t(68)=-15.92$, $p=1.10^{-24}$) and *Distant* ($\beta=-0.10\pm0.02$, $(68)=-6.27$, $p=3.10^{-8}$) conditions, suggesting that it takes more time to provide a rare response compared to a common one (Figure 3A). We also observed that the cue steepness (how strongly connected is the first associate of the cue, see Methods 4.3.1) also significantly shortened response time for *First* responses but not significantly for *Distant* responses ($\beta_{\text{First}}=-0.13\pm0.02$, $t(68)=-8.5$, $p=3.10^{-12}$; $\beta_{\text{Distant}}=-0.02\pm0.01$, $t(68)=-1.16$, $p=0.25$, Figure 3B).

FGAT responses: adequacy and originality
Using adequacy and originality ratings provided by the participants, we found that *First* responses were rated as more adequate than *Distant* responses ($\text{Adequacy}_{\text{First}}=86.47\pm0.99$, $\text{Adequacy}_{\text{Distant}}=77.24\pm1.23$, $t(68)=9.29$, $p=1.10^{-13}$), but *Distant* responses were rated as more original than *First* responses ($\text{Originality}_{\text{First}}=33.80\pm1.74$, $\text{Originality}_{\text{Distant}}=64.43\pm1.37$, $t(68)=-16.36$, $p=3.10^{-25}$). Note that the difference in originality ratings (*First* versus *Distant* responses) was greater than the difference in adequacy ratings ($t(68)=-13.87$, $p=2.10^{-21}$), suggesting that *Distant* responses were found both adequate and original, i.e., creative, while *First* responses were mainly appropriate (Figure 3C).

FGAT responses: likeability
Last, we considered that response time and typing speed could reflect an implicit valuation of responses (45). To test whether an implicit subjective valuation of response happened during the FGAT creative condition (*Distant*), we investigated the link between response time, typing speed, and the likeability of

their own FGAT responses (see Methods 4.3.1). We found that response time in the *Distant* condition decreased with likeability ($\beta_{Distant}$=-0.15±0.02, t(68)=-7.25, p=5.10^{-10}) and that typing speed increased with it ($\beta_{Distant}$=0.08±0.02, t(68)=3.88, p=2.10^{-4}). Participants were faster for providing *Distant* FGAT responses they liked the most. The pattern was different in the *First* condition, in which we observed a significant increase of response time with likeability (β_{First}=0.08±0.02, t(68)=3.78, p=3.10^{-4}) and no significant effect of likeability on typing speed (β_{First}=0.009±0.02, t(68)=0.36, p=0.72). The effects of likeability significantly differed at the group level between the *First* and *Distant* conditions (*Distant* versus *First* effect of likeability on response time: t(68)=-7.30, p=4.10^{-10}; on typing speed: t(68)=2.21, p=0.03, Figure 3D).
Note that the link between likeability rating and response time, or typing speed remains after removing confounding factors (adequacy and originality ratings, SI Table S1).

Together, those findings suggest that likeability might have been cognitively processed during the FGAT task and influenced the behavior, particularly during the FGAT *Distant* condition, which is assumed to require an evaluation of the response before the participants typed their answers. We also found that likeability ratings drove choices (choice task, see SI Supplementary Results and Figure S1), suggesting that likeability is relevant both in the FGAT *Distant* condition, and in binary choices linked to creative response production. We next assessed how likeability ratings relied on adequacy and originality ratings.

2.1.2. Likeability depends on originality and adequacy ratings

To better understand how subjects built their subjective value and assigned a likeability rating to a cue-response association, we focused on the behavior measured during the rating tasks. In the rating tasks, participants judged a series of cue-response associations in terms of their likeability, adequacy and originality (see Figure 2 and Methods 4.2.2). Here, we explored the relationship between those three types of ratings.
We first observed that likeability increased with both originality and adequacy (Figure 4). Then, to precisely capture how adequacy and originality contributed to likeability judgments, we compared 12 different linear and non-linear models (see Methods 4.3.4). Among them, the Constant Elasticity of Substitution (CES) model out performed (41) the alternatives (Estimated model frequency: Ef=0.36, Exceedance probability: Xp=0.87). CES combines originality and adequacy with a weighting parameter α and a convexity parameter δ into a subjective value (likeability rating) (see equation in Figure 1 and fit in Figure 4). Mean values of α and δ are detailed in SI.

Overall, these results indicate that subjective valuation seems to occur during idea generation, as we observed significant relationships between response speed and likeability ratings in the generation task. Additionally, the rating tasks allow us to characterize the *valuator* module as the Constant Elasticity of Substitution utility function (CES), that builds a subjective value from adequacy and originality ratings.

2.2. Computational modelling of the valuator module

The goal of our computational model is to explain and predict the behavior of participants in the FGAT, by modelling an *explorer* that generates a set of candidate ideas, a *valuator* that assigns a subjective value to each candidate idea, and a *selector* that selects a response based (or not) on this subjective value. Our computational model thus needed to be able to predict likeability of any potential cue-response associations, including those that have not been rated by our participants (see section 4.2.2), and those that have not been expressed by participants during the FGAT *Distant* condition (hidden candidate ideas).
We found that adequacy and originality rating could be correctly predicted by associative frequency (see SI Supplementary Results and Figure S2). Adequacy ratings could be well fitted through a linear relation with frequency (Ef_{lin}=0.86, Xp_{lin}=1), and originality could be estimated through a mixture of linear and quadratic link with frequency. This result allows us to estimate adequacy and originality of any cue-response association for a given participant.
Importantly, we explored the validity of the *valuator* module using estimated adequacy and originality. We estimated likeability from the estimated adequacy and originality, using the individual parameters of the CES function mentioned above. We found a strong relationship between estimated and real likeability judgements (mean r=0.24±0.02, t(68)=11.04 p=8.10^{-17}).

This result is not only a critical validation of our model linking likeability, originality and adequacy, but also allows defining a set of parameters for each individual for the *valuator* module. Thanks to that set of parameters, we were able to significantly predict the originality, adequacy, and likeability ratings of any cue-response association based on its objective associative frequency. Henceforth, in the next analyses, likeability, adequacy, and originality estimated through that procedure will be referred to as the "estimated" variables.

In the next section, using computational modelling, we address the second aim of our study, which was to develop the *explorer* and the *selector* and determine which module the *valuator* drives the most.

2.3. Computational modelling of the exploration and selection modules

2.3.1. Model description and overall strategy

As we do not have a direct access to the candidate ideas that participants explored before selecting and producing their response to each cue during the FGAT task, we adopted a computational approach that uses random walk simulations ran on semantic networks (one per FGAT cue) to develop the *explorer* module. We built a model that coupled random walk simulations (*explorer*) to a valuation (*valuator*) and selection (*selector*) function (Figure 1). The model takes as input an FGAT cue and generates responses for the *First* and *Distant* conditions, allowing us to ultimately test how similar the predicted responses from the model were to the real responses of the participants.
In the following analyses, we decompose the model into modules (random walks and selection functions) and investigate by which variable (estimated likeability, estimated originality, estimated adequacy, associative frequency or mixtures) each module is more likely to be driven.

To assess the validity of the model, we developed it and conducted the analyses on 46 subjects (2/3 of them) and then cross-validated the behavioral predictions on the 23 remaining participants.

2.3.2. Modelling the *explorer* module using random walks on semantic networks

For each cue, we built a semantic network from the *Dictaverf* database that was enriched from both *First* and *Distant* FGAT responses from all participants (see Methods 4.3.6). Then, to investigate whether exploration could be driven by likeability, we compared five censored random walks (RW), each with different transition probabilities between nodes (random, associative frequency, adequacy, originality or likeability, see Methods 4.3.6). For each random walk, subject, and cue, we computed the probability of the random walk to visit the *First* and the *Distant* responses nodes (Figure 5A). We found that the frequency-driven random walk (RWF) had the highest chance to walk through the *First* (mean probability $= 0.30 \pm 0.01$; all $p<10^{-33}$) and *Distant* (mean probability $= 0.05 \pm 0.004$; all $p<10^{-4}$) responses. This result suggests that the *explorer* module may be driven by associative frequency between words in semantic memory. According to this result, we pursued the analyses and simulations with the RWF as an *explorer* module for both *First* and *Distant* responses.

2.3.3. Visited nodes with the RWF as a proxy for candidate responses

To define sets of candidate responses that will then be considered as options by the *selector* module, we simulated the RWF model for each subject and each cue over 18 (see Methods 4.2.4 and 4.3.6). Each random walk produced a path: i.e., a list of words (nodes) visited at each iteration. Each node is associated with a rank (position in the path), which will then be used as a proxy of response time. As a sanity check, we compared the list of words obtained from those random walks to the participants' responses to a fluency task on six of the FGAT cues (see Methods 4.2.4). For each subject, we identified the common words between the model path and the fluency responses. Then, using a mixed-effect linear regression with participants and cues as random factors (applied to both intercept and slope), we regressed the node model rank against its corresponding fluency rank. We found a significant fixed effect of the fluency rank ($\beta= 0.12\pm0.03$, $t(649)=3.35$, $p=8.10^{-3}$, SI Figure S3), suggesting that those simulations provide an adequate proxy for semantic memory exploration.
Together, results reported in sections 2.3.2 and 2.3.3 suggest that a censored random walk driven by the frequency of word associations provides a good approximation of semantic exploration during

response generation in the FGAT task and that likeability has a negligible role during that phase. Hence, valuation does not seem to play a significant role in the *explorer* module.

2.3.4. Modelling the *selector* module as a decision function

We then explored the possible factors driving individual decisions to choose a given response (*selector* module) among the word nodes visited by the *explorer* module.
To investigate the selection of *First* and *Distant* responses among all nodes in each path, i.e., on which dimension responses were likely to be selected, we compared seven choice models with different variables as input: random values, node rank (first visited nodes have higher chances of being selected), estimated adequacy, estimated originality, interaction between estimated adequacy and originality, sum of estimated adequacy and originality, and estimated likeability (see Methods 4.3.7). We found that estimated adequacy was the best criterion to explain the selection of *First* responses ($Ef_{adequacy}$=0.89, $Xp_{adequacy}$=1) and likeability was the best criterion to explain the selection of *Distant* responses ($Ef_{likeability}$=0.66, $Xp_{likeability}$=0.99) (Figure 5B). These results indicate that valuation (based on individual likeability) is needed to select a creative response in the creative condition of the FGAT (*Distant*).

2.4. Validity of the full model: does it predict behavioral responses in the test group?

After having characterized the equations and individual parameters of the *valuator* on all participants using the rating tasks, and of the *explorer* and *selector* modules on a subset of participants (n_1=46), we checked whether this model could generate surrogate data similar to the behavior of the remaining participants (test group, n_2=23). We simulated behavioral data and response time from the full model (*explorer*, *valuator*, *selector*), depicted in Figure 1 (See Methods 4.3.8).
We analyzed the behavior of the simulated data the same way we analyzed the behavior of the real human data of the test group (see Methods 4.3.8). We found the same patterns at the group level (SI Table S2, Figure 6 and S4): 1) *First* responses were much more common than *Distant* responses (Figure 6A, B); 2) the rank in path decreased with the group frequency of responses, both for *First* and *Distant* responses (Figure 6A, B), confirming that it takes more time to provide a rare response compared to a common one; 3) Ranks decreased with the cue steepness, both for *First* and *Distant* responses (Figure 6C, D); 4) Ranks of the *Distant* responses decreased with estimated likeability. The effect was significant only for *Distant* responses and the difference between regression estimates for *First* and *Distant* responses was significant. (Figure 6E, F); 5) *First* responses were more appropriate than *Distant* responses, but *Distant* responses were more original than *First* responses. The difference in originality rating between *First* and *Distant* responses was bigger than the difference in adequacy (SI Figure S4).
Additionally, we checked whether the surrogate data generated by the model for each participant was relevant at the inter-individual level. We estimated the *selector* parameters for the test group and conducted the analyses on all participants to increase statistical power. We found that the mean response time per participant across trials of the FGAT *Distant* condition was correlated with the mean rank of *Distant* responses across trials in the model exploration path (r=0.72, p=1.10^{-4}). Similarly, the mean associative frequency (*Dictaverf*) of participants' *Distant* responses was significantly correlated to the mean frequency of the model *Distant* responses (r=0.53, p=9.10^{-3}). These results mean that the model successfully predicted individual behavioral differences in the FGAT task.

2.5. Relevance of model parameters for creative abilities

Finally, to assess the relevance of the individual model parameters in relation to the FGAT task for creative abilities, we defined two sets of variables: FGAT parameters and scores reflecting the *valuator*, *selector* and *explorer* individual characteristics, and Battery scores related to several aspects of creativity (see Methods 4.2.4 and SI Methods). We conducted a canonical correlation analysis between those two sets in all participants and found one canonical variable showing significant dependence between them (r=0.61, p=0.0057). When assessing which variables within each set had the highest coefficient to the canonical score, we found that the two likeability parameters (α and δ, from the *valuator*), the inverse temperature (choice stochasticity, from the choice task, see SI results and Methods 4.3.3) of the *Distant* response selection (from the *selector*) and the *First* response associative frequencies were significantly contributing the FGAT canonical variable. Additionally, fluency score from the fluency task and from the alternative uses task (AUT), creativity self-report, and PrefScore (self-report of preferences regarding ideas) significantly contributed to the Battery canonical variable. No significant contribution was observed from creative activities (C-Act) and achievements (C-Ach) in real life scores (Table S3, Figure 7). Overall, this significant canonical correlation indicates that measures of valuation and selection relate to creative behavior.

3. Discussion

3.1. Summary

In the current study, we investigated the role of valuation based on adequacy and originality in idea generation and creativity. We found that people built subjective values of ideas based on their adequacy and originality that guided their preference and impacted their idea generation during the FGAT task. There was a signature of this value in the response speed of the participants during the FGAT task. Then, we investigated whether preferences were more likely to impact semantic exploration or response selection using a computational model combining random walk on semantic networks (*explorer*), the subjective valuation of candidate responses (*valuator*), and decision for response selection (*selector*). We found that semantic exploration was more likely to be driven by the associative frequency between words, independently of the individual goal (providing the first response that comes to mind vs. providing a creative response). On the contrary, response selection was driven by adequacy for an uncreative goal and by likeability for a creative goal. Critically, we have shown that our computational model is able to predict the main behavioral patterns of human participants solely by using individual preference parameters, estimated from rating tasks. Finally, we confirmed the relevance for creative abilities of the individual parameters computed with our model.

3.2. Preferred associations are produced faster when thinking creatively

Using the FGAT task, previously associated with creative abilities (13), we found that *Distant* responses were overall more original and slower in response time than *First* responses. In addition, response time decreased with steepness (only for *First*) and cue-response associative frequency. Those results are in line with the notion that it takes time to provide an original and rare response (46, 47).
Critically, we identified that the likeability of *Distant* responses was negatively linked to response time and positively linked to typing speed. Interpretation of response time can be challenging as it could reflect the easiness of choice (48), the quantity of effort or control required for action (49), motivation (45), or confidence (50). In any case, this result, surviving correction for potential confounding factors (see Results 2.1.1), represents evidence that subjective valuation of ideas occurs during a creative (hidden) choice. To our knowledge, this is the first time that such a result has been demonstrated. With our computational model, we attempt to provide an explanation of a potential underlying mechanism involving value-based idea selection.

3.3. Subjective valuation of ideas drives the selection of a creative response

The striking novelty our results reveal is the role of the *valuator* module coupled with the *selector* module in idea generation. These modules are directly inspired by the value-based decision-making field of research (24, 51). To make any kind of goal-directed choice, an agent needs to assign a subjective value to items or options at stake, so that they can be compared and one of them can be selected (52). Here, we hypothesized that providing a creative response involves such a goal-directed choice that would logically require the subjective valuation of candidate ideas. After finding a behavioral signature of subjective valuation in response time and typing speed, we have shown that *Distant* response selection among a set of options was best explained by likeability judgments. This pattern was similar to the behavior observed in the choice task, explicitly asking participants to choose the response they would have preferred to give in the FGAT *Distant* condition. Valuation is closely related to motivation process, as it is assumed that subjective values would energize behaviors (53). Previous studies have highlighted the importance of motivation in creativity (54, 55). However, those reports were mainly based on interindividual correlations, while our study brings new evidence for the role of motivation in creativity with a mechanistic approach. Our findings support the hypothesis that the Brain Valuation System is involved in creative thinking and paves the way to later investigate its neural response during creative experimental tasks.

Our study also reveals some of the mechanisms about how individual preferences are built and used to make creative choices. Using the rating tasks and comparing several valuation functions, we identified how originality and adequacy ratings were taken into account to build likeability, and determined preference parameters (relative weight of originality and adequacy and convexity of preference) to predict the subjective likeability of any cue-response association. Subjective likeability relied on subjective adequacy and originality. The identified valuation function linking likeability with adequacy and originality, i.e., the Constant Elasticity of Substitution utility function, has been previously used to explain moral choices or economic choices (56, 57, 41), making it an appropriate candidate for the *valuator* module of our model. Overall, these results indicate that likeability is a relevant measure of the individual values that participants attributed to their ideas, and inform us on how it relies on the combination of originality and adequacy.

The second novelty of our study is to provide a valid full computational model composed of an *explorer*, a *valuator* and a *selector* module. We characterized these modules, and brought an unprecedented mechanistic understanding of creative idea generation. This full model is able to generate surrogate data similar to the real human behavior, both at the group and inter-individual level.

3.4. A computational model that provides a mechanistic explanation of idea generation

The computational model presented in the current study is consistent with previous theoretical frameworks involving two phases in creativity: exploration and evaluation/selection (7–10). The *explorer* module was developed using random walks as it had been successfully done in previous studies to mimic semantic exploration (58). Here, we found that the simulated semantic exploration was driven by associative frequency between words, but was not biased by subjective judgments of likeability, adequacy or originality. This result is consistent with a recent study showing that a random walk applied to a semantic network was sufficient to predict the pattern of responses in a fluency task (38). In that study, the authors also demonstrated that creative abilities were linked to the network structure rather than to the search process, replicating a previous result (30). Here, we could not draw any conclusion about the link between the individual semantic network structure and creative abilities, since we did not have access to individual semantic networks and used the same semantic network for all participants. Nevertheless, we have shown that a frequency biased random walk yielded higher probabilities of reaching real individual responses compared to other biased random walks. This result is consistent with the associative theory of creativity (59), which assumes that creative search is facilitated by semantic memory structure, and with experimental studies linking creativity and semantic network structure (60) or word associations (61). Indeed, the random walks that we compared could be combined in three groups: purely random, structure-driven (frequency-biased) and goal-directed (cue-related adequacy, originality and likeability biased). Here, we found that the structure-driven random walk outperformed the random and goal-directed random walks, providing further evidence that semantic search has a spontaneous, bottom-up component, and with previous studies that used free fluency tasks (60) or word association tasks (61).

Overall, our computational approach does not explore the neural mechanism of creative response generation per se, yet it has several strengths. First, it considers creativity as a plural mechanism (three modules) occurring in each individual. Second, it adds to previous research a new framework to explore creativity by combining semantic search and value-based response selection. Third, it allows behavioral predictions at the individual and group level. Classically, creativity is investigated as an ability varying across individuals, and differences between low and high creative abilities are investigated. Although this approach has allowed the discovery of key results about human creativity - such as the importance of semantic network structure or the impact of motivation - it prevents understanding how the human brain implements idea generation and selection, independently of its creative performance. Computational cognitive modelling is now widely used in cognitive neuroscience but it has rarely been applied to neuroscience of creativity. The use of model fitting procedure, model selection and surrogate data generation, in accordance with guidelines suggested by previous work (62), has a high potential for better understanding underlying mechanisms of creativity as demonstrated in our study.

3.5. Limitations

Some limitations to this study need to be acknowledged. First, the present study assesses creative cognition in the semantic domain. To fully validate our computational model and the core role of preference-based idea selection, it is necessary to apply similar analyses on other domains such as drawing or music. Second, to build our model, we made many assumptions, such as the structure of semantic networks, and each of them should be tested explicitly in future studies. Third, our main result concludes on the role of motivation and preferences in idea selection, but their role in the exploration process per se remain to be further understood. Fourth, our model is for now quite serial and needs more development. For instance, the number of ideas considered at each step was fixed in our model, but one should also consider that a smaller number of ideas are evaluated at each step and that the whole process is restarted if a threshold value is not reached. Thus, our model will need some further extension, notably by adding iterations and a "stop" criterion.

3.6. Conclusion

The present study reveals the role of individual preferences and decision making in creativity, by decomposing and characterizing the exploration and the evaluation/selection processes of idea generation. Our findings demonstrate that the exploration process relied on associative thinking while the selection process depended on the valuation of ideas. We also show how preferences are formed by weighting adequacy and originality of ideas. By assessing creativity at the group level, beyond the classical interindividual assessment of creative abilities, the current study paves the way to a new framework for creativity research and places creativity as a complex goal-directed behavior driven by reward signals. Future neuroimaging studies will examine the neural validity of our model.

4. Materials and Methods

4.1. Participants

The study was approved by an official ethics committee. Seventy-one participants were recruited and tested thanks to the PRISME platform of the Paris Brain Institute (ICM). They gave informed consent, and were compensated for their participation. Inclusion criteria were: being right-handed, native French speakers, between 22 and 40 years old, with correct or corrected vision and no history of neurological or psychiatric disease. Two participants were excluded because of a misunderstanding of the instructions, bringing the final number of participants to 69 (41 females and 28 males; mean age: 25.8±4.5; mean level of education: number of study years following French A-levels: 5.0±1.6).

4.2. Experimental design

Each participant performed three types of tasks of creative generation and evaluation of ideas, which were followed by a battery of tests classically used in the laboratory and assessing the participant's creative abilities. All tasks and tests were computerized and administered in the same fixed order for all participants.

4.2.1. Free Generation of Associations Task (FGAT)

The Free Generation of Associations Task (hereafter referred to as FGAT) is a word association task, previously shown to capture aspects of creativity (13) (63). It is composed of two conditions, presented successively, always in the same order. Cue words selection is detailed in SI.

FGAT-first condition

After a 5-trials training session, participants performed the 62 trials of the first condition block (hereafter referred to as FGAT-first). They were presented with a cue word and instructed to provide the first word that came to mind after reading the cue word. They had 10 seconds to find a word and press the spacebar and then were allowed 10 seconds maximum to type it on a keyboard. This condition was used to explore the participants' spontaneous semantic associations and served as a control condition that is not a creative task per se.

FGAT-distant condition

In a different following block, participants were administered 62 trials of the second condition of the task (hereafter referred to as FGAT-distant). On each trial, they were presented with a cue word as in the previous condition and instructed to press the spacebar once they had thought of a word unusually associated with the cue. They were asked to find a distant but understandable associate and to think creatively. They had 20 seconds to think of a word and press the spacebar and then were allowed 10 seconds maximum to type it. This condition was used to measure the participants' ability to produce remote and creative associations intentionally.

4.2.2. Rating tasks

After the FGAT task, participants performed two rating tasks. In the first block, they had to rate how much they liked an association of two words (likeability rating task). Then, in a separate block performed after the Choice task (see below), they had to rate the originality and the adequacy (originality and adequacy rating task) of the same associations as in the likeability rating task.

Likeability Rating task

After a 5-trial training session, participants performed 197 trials in which they were presented with an association of two words (cue-response, see below) and asked to rate how much they liked this cue-response association in a creative context, i.e., how much they like it or would have liked to find it during the FGAT *Distant* condition. A cue-response association was displayed on the screen, and 0.3 to 0.6 seconds later, a rating scale appeared underneath it. The rating scale's low to high values were represented from left to right, without any numerical values but with 101 steps and a segment indicating the middle of the scale (later converted in ratings ranging between 0 and 100). Participants entered their rating by pressing the left and right arrows on the keyboard to move a slider across the rating scale, with the instruction to use the whole scale. Once satisfied with the position of the slider, they pressed the spacebar to validate their rating and went on to the subsequent trial. No time limit was applied, but participants had the instruction to respond as spontaneously as possible. A symbol (a heart for likeability ratings) was placed underneath the scale as a reminder of the dimension on which the words were to be rated.

Originality and Adequacy Ratings

Another Rating task was performed after the Choice task. After a 5-trial training session, participants performed a block of 197 trials. They were asked to rate the same set of associations as in the likeability task, but this time in terms of originality and adequacy, and in a different random order. In the instructions, an original association was described as 'original, unusual, surprising'. An adequate association was described as 'appropriate, understandable meaning, relevant, suitable'. Note that the instructions were given in French to the participants and the adjectives used in here are the closest translation we could find.

For each cue-response association, participants had to rate originality and adequacy dimensions one after the other, in a balanced order (in half of the trials, participants were asked to rate the association's adequacy before its originality, and in the other half of the trials, it was the opposite). The order was unpredictable for the participant. Similar to the likeability ratings, the rating scale appeared underneath the association after 0.3 to 0.6 seconds, with a different symbol below it: a star for originality ratings and a target for adequacy ratings, as depicted in Figure 2.

Cue-word associations

The 197 cue-response associations presented in the rating tasks and choice task were built with 35 FGAT cue words randomly selected for each participant, at the end of FGAT with a MatLab script that implemented an adaptive design with the following rules. Each cue word was associated with seven words, amounting to 245 possible associations in total. The seven associated words for each cue word were selected from the participant's answers and from another dataset collected previously in the lab that gathers the responses of 96 independent and healthy participants on a similar FGAT task (See SI Supplementary Methods for a full description).

4.2.3. Choice task

Between the likeability rating task and the adequacy-originality rating task, participants performed a binary choice task. They had to choose between two words the one they preferred to be associated with a cue in a creative context, i.e., in the FGAT *Distant* context. Instructions were as follows: 'For example, would you have preferred to answer "silver" or "jewellery" to "necklace" when generating original associations during the previous task?' (There was additionally a reminder of the FGAT *Distant* condition, in the instructions). Details of the task can be found in SI Supplementary Methods.

4.2.4. Battery of creativity tests

A battery of creativity tests run on Qualtrics followed the previous tasks, in order to assess creative abilities and behavior of the participants. It was composed of the alternative uses task (AUT), the inventory of Creative Activities and Achievements (ICAA), a self-report of creative abilities, a scale of preferences in creativity between adequacy and originality (SPC) and a fluency task on six FGAT cues. There are described in detail in the Supplementary Methods.

4.3. Statistical analysis and computational modelling

All analyses were performed using Matlab (MATLAB. (2020). 9.9.0.1495850 (R2020b). Natick, Massachusetts: The MathWorks Inc.). Model fitting and comparison were conducted using the VBA toolbox (https://mbb-team.github.io/VBA-toolbox/) (65).

4.3.1. Analyses of the FGAT responses

The main behavioral measures of interest in the FGAT task are the response time (pressing the space key to provide an answer), the typing speed (number of letters per second), and the associative frequency of the responses. This frequency was computed based on a French database called *Dictaverf* (http://dictaverf.nsu.ru/)(66) built on spontaneous associations provided by at least 400 individuals in response to 1081 words (each person saw 100 random words). Frequencies were log-transformed to take into account their skewed distribution toward 0. Cues varied in terms of steepness (the ratio between the associative frequency of the first and distant associate of a given cue word), which also constituted a variable of interest. The ratings provided by subjects on their own responses (adequacy, originality, and likeability) were also used as variables of interest.
Linear regressions were conducted at the subject level between normalized variables. Significance was tested at the group level using one sample two-tailed t-test on coefficient estimates.

4.3.2. Model fitting and comparison

Every model/module was fitted at the individual level to ratings and choices using the Matlab VBA-toolbox, which implements Variational Bayesian analysis under the Laplace approximation (67, 68). This iterative algorithm provides a free-energy approximation to the marginal likelihood or model evidence, which represents a natural trade-off between model accuracy (goodness of fit) and complexity (degrees of freedom) (69, 70). Additionally, the algorithm provides an estimate of the posterior density over the model free parameters, starting with Gaussian priors. Individual log-model evidence were then taken to group-level random-effect Bayesian model selection (RFX-BMS) procedure (68, 71). RFX-BMS provides an exceedance probability (Xp) that measures how likely it is that a given model (or family of models) is more frequently implemented, relative to all the others considered in the model space, in the population from which participants were drawn (68, 71).

We conducted the first model comparison to determine which variable (Adequacy A, Originality O or Likeability L) best explained choices (Methods 4.3.3). The second model comparison was performed to identify which utility function (*valuator* module) best explained how originality and adequacy were combined to compute likeability (Methods 4.3.4). The third one aimed at establishing relationships between adequacy and originality ratings and associative frequency of cue and responses (Methods 4.3.4). The fourth one aimed at identifying the best possible input variable for the *selector* module (Methods 4.3.7).

4.3.3. Relationship between choices and ratings

Logistic regression was applied to choices as a dependant variable, with likeability (L), originality (O) or adequacy (A) ratings as regressors. Choices were analyzed at the subject level and tested for significance at the group level (random-effect analysis) using two-tailed, paired, Student's t-tests. The softmax function used to determine which variable (V) among likeability, adequacy or originality ratings was better explaining the proportion choices for left options (P(Left)) against right options is the following:

$$P(Left) = \frac{1}{1 + e^{-\frac{V_{Left} - V_{Right} - d}{\beta_{choice}}}}$$

With *d* being a constant term aiming at capturing any bias towards one side and β_{choice} the temperature (choice stochasticity).

4.3.4. *Valuator* module: combining likeability originality and adequacy of the rating tasks with responses associative frequency

The ratings were used to estimate the likeability of a given response to a cue from its adequacy and originality, themselves estimated from its associative frequency.

Likeability ratings relationship with adequacy and originality ratings
First, we fitted 12 different functions to likeability ratings capturing linearly (or not) the relationship between likeability (L) and adequacy (A) and originality (O):

- Linear models:

$$L_i = \beta A_i \qquad L_i = \alpha O_i + (1-\alpha)A_i \qquad L_i = \alpha O_i + \beta A_i$$

- Linear with interaction term models:

$$L_i = \alpha O_i + (1-\alpha)A_i + \gamma O_i * A_i \qquad L_i = \alpha O_i + \beta A_i + \gamma O_i * A_i \qquad L_i = \gamma O_i * A_i$$

- Non-linear models (with the same non-linearity on both dimensions):

$$L_i = (\alpha {O_i}^{\delta} + (1-\alpha){A_i}^{\delta})^{\frac{1}{\delta}}\ (CES) \qquad L_i = (\alpha {O_i}^{\delta} + \beta {A_i}^{\delta})^{\frac{1}{\delta}} \qquad L_i = \alpha {O_i}^{\delta} + \beta {A_i}^{\delta}$$

  The first non-linear model is also referred as Constant Elasticity of Substitution (CES) (57)

- Non-linear models (with different non-linearity on both dimensions):

$$L_i = \beta {A_i}^{\delta} \qquad L_i = \alpha {O_i}^{\delta} + (1-\alpha){A_i}^{\varepsilon} \qquad L_i = \alpha {O_i}^{\delta} + \beta {A_i}^{\varepsilon}$$

Greek letters correspond to free parameters estimated with the fitting procedure described below; i refers to a given cue-response association.

Adequacy and originality ratings relationship with associative frequency
Second, we investigated how adequacy and originality were linked to associative frequency between a cue and a response F_{ci}. For each dimension X (A or O), we compared three models:

$$X_i = \mu_X^l\, log(F_{ci}) \qquad X_i = \mu_X^q\, log(Fc_i)^2 \qquad X_i = \mu_X^l\, log(F_{ci}) + \mu_X^q\, log(Fc_i)^2$$

μ_X^l corresponds to the linear regression coefficient and μ_X^q to the quadratic regression coefficient.

4.3.5. Model identification group and test group

We randomly split our group of participants into two subgroups, one group to develop the *explorer* and *selector* modules (2/3 of the group: 46 subjects) and one group to validate the full module (combination of the *explorer*, *valuator* and *selector* modules) by comparing its behavioral prediction to the actual behavior of the participants (23 subjects).

4.3.6. Modelling the explorer module

Construction of semantic networks

The *Dictaverf* database consists of 1081 cue words associated with 23340 other words and is organized as a matrix M of i rows and j columns, with associative frequencies directed from the cue-words i to the response-words in j. We used this database combined with FGAT responses to build a symmetric adjacency matrix C of word associations for each cue word, applying the following procedure.

1) A list of all FGAT responses (*First* and *Distant*) from the current and the former datasets of the lab was created for each cue.
2) Then, for each response in the list:
   a. If it was already associated with the cue in *Dictaverf*, for example, "learning" in response to "school", then: C(school,learning) = C(learning,school) = M(school,learning).
   b. If it was not associated with the cue in *Dictaverf*: we looked for it in the whole database and identified all potential intermediate nodes between the cue and the word (any other words associated with both the cue and the response). For instance, one subject responded "anxiety" to "school". "Anxiety" was not directly linked to "school" in M, but it was connected to "studies", which was connected to "school". Then, "anxiety" was added as a row (and column for symmetry) in the matrix C, and frequency between "Anxiety" and "studies" was defined as the frequency between "studies" and "anxiety" from M. "Studies" was also added as a row (and column) in C and the frequency between "studies" and "school" was set as the frequency between "school" and "studies":
   C(anxiety, studies)=C(studies, anxiety)=M(studies, anxiety)
   and
   C(studies, school)=C(school, studies)=M(studies, school).
   In this example, when then building a network based on C (see below), "anxiety" is thus connected to "school" via the node "studies".
   This procedure was applied to all potential intermediate nodes, independently of the number of intermediates.

This procedure yielded 62 symmetric C matrices (one per cue) with a size of around 1022 by 1022 words (ranging between 689 and 1186). The first row and column of each matrix correspond to the cue on which the matrix has been built (SI Figure S5 for visual explanation).

Sixty-two networks N were built based on those C matrices as unweighted and undirected graphs (two nodes were linked by an edge if the frequency of association between them was higher than 0).

Random walks variants and implementation

We used censored random walks that start at a given cue and walk within its associated network N. Censored random walks have the property to not return to previously visited nodes. In case of a dead-end, the censored random walk starts over from the cue but does not go back to previously visited nodes. The five following variants of censored random walks were applied to the semantic networks to simulate potential paths.

- The random walk random (RWR) was a censored random walk starting at the cue and with uniform distribution of probabilities of transition from the current node to each of its neighbours (excluding previously visited nodes).
- The random walk frequency (RWF) was a censored random walk biased by the associative frequency between nodes, where the probability of transition from one node to another one is defined as follows:

$$P_{ij}^{F} = \frac{F_{ij}}{\sum_j F_{ij}}$$

with P_{ij}^{F} the probability of transition to node j, F_{ij} the frequency of the association in the C matrices described above with the current node i, and j all the other nodes linked to the current node n.

- Three additional censored random walks were run. They were biased by adequacy (RWA), originality (RWO), or likeability (RWL) of association between nodes and cue, where the probability of transition from one node to another one is defined as follows:

$$P_{ij}^X = \frac{X_{cj}}{\sum_j X_{cj}}$$

with P_{ij}^X the probability of transition from node i to node j, X_{cj} the estimated adequacy, originality, or likeability of the node j with the cue node c, j are all the nodes linked to the current node i.

Estimated adequacy, originality and likeability of all the network nodes (X_i) were computed based on the results of the model comparison performed in the first section (see Methods 4.3.4). The following equations were consequently used:

$$A_i = \mu_A log\ (F_{ci})$$

$$O_i = \mu_O^l\ log(F_{ci}) + \mu_O^q\ log(F_{ci})^2$$

$$L_i = (\alpha {O_i}^\delta + (1-\alpha){A_i}^\delta)^{\frac{1}{\delta}}$$

With F_{ci} as the frequency of association between the node and the cue.
Note that if a given node was not directly linked to the cue, we computed F_{ci} as the cumulative product of the frequency of association of the nodes belonging to the shortest path between the cue and the node. For example: $F_{school\text{-}anxiety} = F_{school\text{-}studies} \times F_{studies\text{-}anxiety}$.
Also, if one node (cue-word association) has actually been rated by the participant, μ_A, μ_O^l, μ_O^q, α, and δ were estimated without that particular cue-word association (leave-one-out procedure) to avoid double-dipping. For example, if the cue-response "School-Anxiety" was rated at trial t by a participant, the predicted adequacy, originality, and likeability of trial t for that participant were computed with parameters estimated with all trials except trial t. This procedure lengthens the processing time of the random walks RWA, RWO and RWL but has the advantage of avoiding double-dipping.

The number of steps performed by each random walk was constant across cues and participants and was defined by the median of fluency score among the group, i.e., 18 steps, resulting in no more than 17 visited nodes.

Probability of reaching First and Distant responses for each participant and cue
We computed the probability of reaching the *First* and *Distant* responses (Targets T) from a starting node cue (c) for each type of random walk as follows:

$$P_{c,T=} \sum_{G=a}^{z} \prod_{\substack{i=c \\ i,j \in G}}^{j=T} P_{i,j}$$

With G representing all possible paths between c and T, ranging from the shortest one (a) to the longest one (z) (limited to 18 steps) and i and j all pairs of nodes belonging to each path, linked by a transition probability $P_{i,j}$. In other words, it corresponds to the sum of the cumulative product of edge weights for all the possible paths between the cue and the target shorter than 18 steps.

4.3.7. Decision functions as the *selector* module

Next, we intended to decipher the criteria determining the selection of a given response.
For each subject and cue, we simulated RWF as described above and retained the paths that contained both the *First* and *Distant* response of the subject for further analyses (the number of excluded cues ranged between 0 and 31 trials over 62, M=9.04 trials, exclusion mainly due to missing responses from participants either in the FGAT *First* or *Distant* condition).
For each subject, we built two response matrices R^F and R^D of the same size *n* by *t, t* being the number of cues (equivalent of trials within one FGAT condition) retained (53 cues per subject on average) and *n* the number of nodes visited by the RWF (fixed at 18) (See SI Figure S6 for visual explanation). Those matrices were filled with zeros, except for nodes and trials that corresponded to the actual participant response. R^F contains ones for cells actually corresponding to the subject's *First* response (one 1 per column), and R^D contains ones for cells corresponding to the subject's *Distant* response. In order to determine the variable on which the selector module was likely to rely on, we built and compared seven matrices of values M_x of size *n* by *t*:

- M_r: random values in the matrix
- M_P: matrix with value decreasing with the order in the path
- M_A: matrix with estimated adequacy of each visited node in the path
- M_O: matrix with estimated originality of each visited node in the path
- M_L: matrix with estimated likeability of each visited node in the path
- M_{A+O}: Sum of A and O
- M_{A*O}: Product of A and O

M_{A+O} and M_{A*O} were added as controls for likeability, which relies on a non-linear weighted sum of adequacy and originality (CES).
Using the VBA toolbox, we fitted the following multivariate *softmax* functions to R^F and R^D separately for the seven different matrices:

$$P(R_{i,t}^F) = \frac{e^{-X_{i,t}/\beta^F}}{\sum_{k=1}^{n} e^{(-X_{k,t}/\beta^F)}} \qquad P(R_{i,t}^D) = \frac{e^{-(X_{i,t})/\beta^D}}{\sum_{k=1}^{n} e^{(-(X_{k,t})/\beta^D)}}$$

P is the probability of the node *i* to be selected as a response (R) *First* (F) or *Distant* (D) among all the possible nodes *k* belonging to the *n* options from the paths at trial t (for a given cue). X corresponds to the values within the seven different input matrices. β^F and β^D are free parameters estimated per subject, corresponding to the temperature (choice stochasticity).
We then compared the seven models for the *First* and *Distant* response separately and reported the results of the model comparison in the results.

4.3.8. Cross-validation of the model: comparing the surrogate data to human behavior

To simulate the behavior of the remaining 23 subjects, we combined all the previously described modules together and released some constraints imposed by the model investigation. We applied RWF with 18 steps on the built networks (see Methods 4.3.6) and assigned values to each visited node according to each subject's valuator module parameters. The list of visited nodes (candidate responses) for each cue and each subject was simulated without the constraint of containing participants' *First* and *Distant* responses. The selection was made using an *argmax* rule on adequacy (winning value for the selector module) for the *First* response and on likeability (winning value for the selector module) for the *Distant* response (as we do not have the selection temperature parameters β^F and β^D for those remaining subjects). We ran 100 simulations per individual following that procedure.
The rank in the path was used as a proxy for response time, and we analyzed surrogate data in the exact same way as subject behavior.
For statistical assessment, regression estimates of ranks against frequency, steepness, and estimated likeability were averaged across 100 simulations per individuals, and significance was addressed at the group level (one representative simulation was used in Figures 6 and S4). For this analysis, group frequency of response was computed instead of *Dictaverf* associative frequency to 1) avoid any confounds with the structure of the graph, built with *Dictaverf*, and 2) compare the distribution of frequencies relative to the group.

4.3.9. Canonical correlation

To investigate the link between creative abilities and our task and model parameters, we extracted the individual task scores and model parameters and grouped them together, into the labelled "FGAT scores and parameters". We grouped the scores obtained from the battery of creativity test and labelled them "Battery scores". We conducted a canonical correlation between those two sets of variables and checked for significance of correlation between the computed canonical variables of each set. Note that a canonical correlation analysis can be compared to a Principal Component Analysis, in the sense that common variance between two data sets is extracted into canonical variables (equivalent of principal components). Canonical variables extracted for each data set are ordered in terms of strength of correlations between the two data sets. Each variable within a data set has a loading coefficient that indicates its contribution to the canonical variable. Here, we extracted the coefficients of each variable on its respective canonical variable and reported them.

5. Acknowledgements

EV is funded by the 'Agence Nationale de la Recherche' [grant numbers ANR-19-CE37-001-01]. The research also received funding from the program 'Investissements d'avenir' ANR-10- IAIHU-06. MOT is funded by Becas-Chile of ANID. ALP was supported by the «Fondation des Treilles». The study was funded by the Paris Region Fellowship Program, "Horizon 2020 Marie Skłodowska- Curie n° 945298. The overall project to which this study belongs has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 101026191. We thank all the participants to the study, Tess Brogard who helped collecting the data, Mehdi Khamassi who provided advices on the model structure.

6. Data and code availability

Data and code will be made available upon publication.

7. References

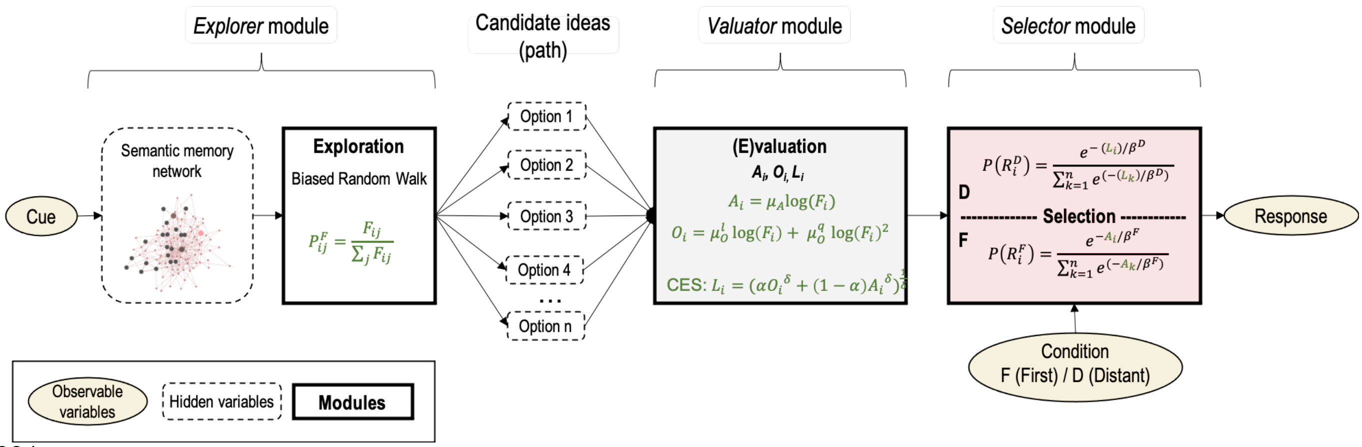
1. A. Dietrich, The cognitive neuroscience of creativity. *Psychon. Bull. Rev.* **11**, 1011–1026 (2004).
2. M. A. Runco, G. J. Jaeger, The Standard Definition of Creativity. *Creat. Res. J.* **24**, 92–96 (2012).
3. R. E. Jung, O. Vartanian, *The Cambridge Handbook of the Neuroscience of Creativity* (Cambridge University Press, 2018).
4. A. Dietrich, R. Kanso, A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychol. Bull.* **136**, 822–848 (2010).
5. E. Volle, "Associative and controlled cognition in divergent thinking: Theoretical, experimental, neuroimaging evidence, and new directions" in *The Cambridge Handbook of the Neuroscience of Creativity*, Cambridge Handbooks in Psychology., Cambridge University Press, (Editors: R.E. Jung and O. Vartanian, 2017).
6. P. T. Sowden, A. Pringle, L. Gabora, The shifting sands of creative thinking: Connections to dual-process theory. *Think. Reason.* **21**, 40–60 (2015).
7. V. Mekern, B. Hommel, Z. Sjoerds, Computational models of creativity: a review of single-process and multi-process recent approaches to demystify creative cognition. *Curr. Opin. Behav. Sci.* **27**, 47–54 (2019).
8. O. M. Kleinmintz, T. Ivancovsky, S. G. Shamay-Tsoory, The two-fold model of creativity: the neural underpinnings of the generation and evaluation of creative ideas. *Curr. Opin. Behav. Sci.* **27**, 131–138 (2019).
9. D. T. Campbell, Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychol. Rev.* **67**, 380 (1960).
10. D. K. Simonton, Donald Campbell's Model of the Creative Process: Creativity as Blind Variation and Selective Retention. *J. Creat. Behav.* **32**, 153–158 (1998).
11. M. Ellamil, C. Dobson, M. Beeman, K. Christoff, Evaluative and generative modes of thought during the creative process. *NeuroImage* **59**, 1783–1794 (2012).
12. R. E. Beaty, M. Benedek, P. J. Silvia, D. L. Schacter, Creative Cognition and Brain Network Dynamics. *Trends Cogn. Sci.* **20**, 87–95 (2016).
13. D. Bendetowicz, *et al.*, Two critical brain networks for generation and combination of remote associations. *Brain* (2017) https:/doi.org/10.1093/brain/awx294 (December 7, 2017).
14. M. Donzallaz, J. M. Haaf, C. Stevenson, Creative or Not? Hierarchical Diffusion Modeling of the Creative Evaluation Process (2021) https:/doi.org/10.31234/osf.io/5eryv (April 8, 2021).
15. N. Mayseless, J. Aharon-Peretz, S. Shamay-Tsoory, Unleashing creativity: The role of left temporoparietal regions in evaluating and inhibiting the generation of creative ideas. *Neuropsychologia* **64**, 157–168 (2014).
16. F. Huang, J. Fan, J. Luo, The neural basis of novelty and appropriateness in processing of creative chunk decomposition. *NeuroImage* **113**, 122–132 (2015).
17. J. Ren, *et al.*, The function of the hippocampus and middle temporal gyrus in forming new associations and concepts during the processing of novelty and usefulness features in creative designs. *NeuroImage* **214**, 116751 (2020).
18. K. Rataj, D. S. Nazareth, F. van der Velde, Use a Spoon as a Spade?: Changes in the Upper and Lower Alpha Bands in Evaluating Alternate Object Use. *Front. Psychol.* **9** (2018).
19. C. Rominger, *et al.*, Functional coupling of brain networks during creative idea generation and elaboration in the figural domain. *NeuroImage* **207**, 116395 (2020).
20. M. F. S. Rushworth, T. E. J. Behrens, Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* **11**, 389–397 (2008).

21. A. D. Redish, N. W. Schultheiss, E. C. Carter, The Computational Complexity of Valuation and Motivational Forces in Decision-Making Processes. *Curr. Top. Behav. Neurosci.* **27**, 313–333 (2016).
22. A. Shenhav, U. R. Karmarkar, Dissociable components of the reward circuit are involved in appraisal versus choice. *Sci. Rep.* **9**, 1958 (2019).
23. M. Lebreton, S. Jorge, V. Michel, B. Thirion, M. Pessiglione, An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging. *Neuron* **64**, 431–439 (2009).
24. D. J. Levy, P. W. Glimcher, The root of all value: a neural common currency for choice. *Curr. Opin. Neurobiol.* (2012) https:/doi.org/10.1016/j.conb.2012.06.001 (October 22, 2012).
25. A. Lopez-Persem, P. Domenech, M. Pessiglione, How prior preferences determine decision-making frames and biases in the human brain. *eLife* **5**, e20317 (2016).
26. T. A. Hare, C. F. Camerer, A. Rangel, Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System. *Science* **324**, 646–648 (2009).
27. P. Domenech, J. Redouté, E. Koechlin, J.-C. Dreher, The Neuro-Computational Architecture of Value-Based Selection in the Human Brain. *Cereb. Cortex* **28**, 585–601 (2018).
28. S. A. Chermahini, B. Hommel, The (b)link between creativity and dopamine: Spontaneous eye blink rates predict and dissociate divergent and convergent thinking. *Cognition* **115**, 458–465 (2010).
29. M. Pessiglione, F. Vinckier, S. Bouret, J. Daunizeau, R. Le Bouc, Why not try harder? Computational approach to motivation deficits in neuro-psychiatric diseases. *Brain* **141**, 629–650 (2018).
30. Y. N. Kenett, D. Anaki, M. Faust, Investigating the structure of semantic networks in low and high creative persons. *Front. Hum. Neurosci.* **8** (2014).
31. M. Benedek, T. Könen, A. C. Neubauer, Associative abilities underlying creativity. *Psychol. Aesthet. Creat. Arts* **6**, 273 (2012).
32. M. Benedek, *et al.*, How semantic memory structure and intelligence contribute to creative thought: a network science approach. *Think. Reason.* **23**, 158–183 (2017).
33. M. Bernard, Y. N. Kenett, M. O. Tellez, M. Benedek, E. Volle, Building individual semantic networks and exploring their relationships with creativity. in *CogSci*, (2019), pp. 138–144.
34. T. Bieth, *et al.*, Dynamic changes in semantic memory structure support successful problem-solving (2021).
35. M. Ovando-Tellez, *et al.*, Brain connectivity-based prediction of real-life creativity is mediated by semantic memory structure. *bioRxiv* (2021).
36. J. C. Zemla, J. L. Austerweil, Modeling Semantic Fluency Data as Search on a Semantic Network. *CogSci Annu. Conf. Cogn. Sci. Soc. Cogn. Sci. Soc. US Conf.* **2017**, 3646–3651 (2017).
37. J. T. Abbott, J. L. Austerweil, T. L. Griffiths, Random walks on semantic networks can resemble optimal foraging. in *Neural Information Processing Systems Conference; A Preliminary Version of This Work Was Presented at the Aforementined Conference.*, (American Psychological Association, 2015), p. 558.
38. Y. N. Kenett, J. L. Austerweil, Examining Search Processes in Low and High Creative Individuals with Random Walks. in *CogSci*, (2016), pp. 313–318.
39. P. A. Samuelson, The numerical representation of ordered classifications and the concept of utility. *Rev. Econ. Stud.* **6**, 65–70 (1938).
40. D. Von Winterfeldt, G. W. Fischer, Multi-attribute utility theory: models and assessment procedures. *Util. Probab. Hum. Decis. Mak.*, 47–85 (1975).
41. A. Lopez-Persem, L. Rigoux, S. Bourgeois-Gironde, J. Daunizeau, M. Pessiglione, Choose, rate or squeeze: Comparison of economic value functions elicited by different behavioral tasks. *PLOS Comput. Biol.* **13**, e1005848 (2017).
42. R. D. Luce, On the possible psychophysical laws. *Psychol. Rev.* **66**, 81 (1959).
43. R. Ratcliff, G. McKoon, The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* **20**, 873–922 (2008).
44. X.-J. Wang, Decision Making in Recurrent Neuronal Circuits. *Neuron* **60**, 215–234 (2008).
45. Y. Niv, Cost, Benefit, Tonic, Phasic: What Do Response Rates Tell Us about Dopamine and Motivation? *Ann. N. Y. Acad. Sci.* **1104**, 357–376 (2007).
46. P. R. Christensen, J. P. Guilford, R. C. Wilson, Relations of creative responses to working time and instructions. *J. Exp. Psychol.* **53**, 82–88 (1957).
47. R. E. Beaty, P. J. Silvia, Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychol. Aesthet. Creat. Arts* **6**, 309–319 (2012).
48. R. Ratcliff, J. N. Rouder, Modeling Response Times for Two-Choice Decisions. *Psychol. Sci.* **9**, 347–356 (1998).
49. M. M. Botvinick, T. S. Braver, D. M. Barch, C. S. Carter, J. D. Cohen, Conflict monitoring and cognitive control. *Psychol. Rev.* **108**, 624 (2001).
50. R. Ratcliff, J. J. Starns, Modeling confidence and response time in recognition memory. *Psychol. Rev.* **116**, 59–83 (2009).

51. A. Lopez-Persem, *et al.*, Four core properties of the human brain valuation system demonstrated in intracranial signals. *Nat. Neurosci.* **23**, 664–675 (2020).
52. A. Rangel, C. Camerer, P. R. Montague, A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* **9**, 545–556 (2008).
53. M. Pessiglione, *et al.*, How the Brain Translates Money into Force: A Neuroimaging Study of Subliminal Motivation. *Science* **316**, 904–906 (2007).
54. M. A. Collins, T. M. Amabile, Motivation and creativity. (1999).
55. C. Fischer, C. P. Malycha, E. Schafmann, The Influence of Intrinsic Motivation and Synergistic Extrinsic Motivators on Creativity and Innovation. *Front. Psychol.* **10**, 137 (2019).
56. P. S. Armington, A Theory of Demand for Products Distinguished by Place of Production. *Staff Pap.* **16**, 159–178 (1969).
57. J. Andreoni, J. Miller, Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica* **70**, 737–753 (2003).
58. J. L. Austerweil, J. T. Abbott, T. L. Griffiths, Human memory search as a random walk in a semantic network. 9.
59. S. Mednick, The associative basis of the creative process. *Psychol. Rev.* **69**, 220–232 (1962).
60. M. Benedek, J. Jurisch, K. Koschutnig, A. Fink, R. E. Beaty, Elements of creative thought: Investigating the cognitive and neural correlates of association and bi-association processes. *NeuroImage* **210**, 116586 (2020).
61. T. R. Marron, *et al.*, Chain Free Association, Creativity, and the Default Mode Network. *Neuropsychologia* (2018) https:/doi.org/10.1016/j.neuropsychologia.2018.03.018 (March 20, 2018).
62. S. Palminteri, V. Wyart, E. Koechlin, The Importance of Falsification in Computational Cognitive Modeling. *Trends Cogn. Sci.* **21**, 425–433 (2017).
63. R. Prabhakaran, A. E. Green, J. R. Gray, Thin slices of creativity: Using single-word utterances to assess creative cognition. *Behav. Res. Methods* **46**, 641–659 (2014).
64. J. Diedrich, *et al.*, Assessment of Real-Life Creativity: The Inventory of Creative Activities and Achievements (ICAA). *Psychol. Aesthet. Creat. Arts* (2017) https:/doi.org/10.1037/aca0000137 (September 7, 2017).
65. J. Daunizeau, V. Adam, L. Rigoux, VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLoS Comput. Biol.* **10**, e1003441 (2014).
66. M. Debrenne, Le dictionnaire des associations verbales du français et ses applications. *Variétés Var. Formes Fr. Palaiseau Éditions L'Ecole Polytech.* **355**, 366 (2011).
67. J. Daunizeau, K. J. Friston, S. J. Kiebel, Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models. *Phys. Nonlinear Phenom.* **238**, 2089–2118 (2009).
68. K. E. Stephan, W. D. Penny, J. Daunizeau, R. J. Moran, K. J. Friston, Bayesian model selection for group studies. *NeuroImage* **46**, 1004–1017 (2009).
69. K. Friston, J. Mattout, N. Trujillo-Barreto, J. Ashburner, W. Penny, Variational free energy and the Laplace approximation. *NeuroImage* **34**, 220–234 (2007).
70. W. D. Penny, Comparing dynamic causal models using AIC, BIC and free energy. *Neuroimage* **59**, 319–330 (2012).
71. L. Rigoux, K. E. Stephan, K. J. Friston, J. Daunizeau, Bayesian model selection for group studies — Revisited. *NeuroImage* **84**, 971–985 (2014).

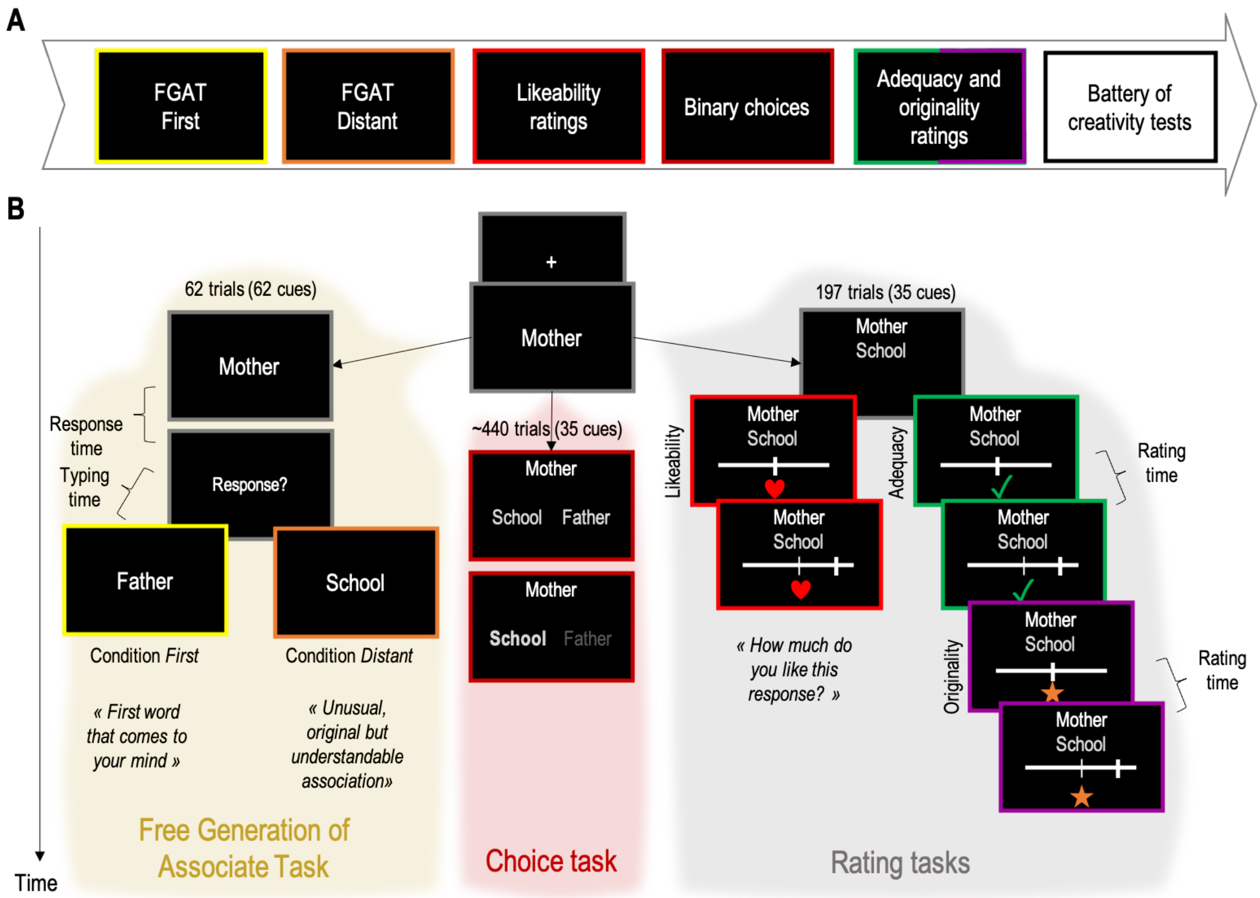
Figures

Figure 1. Schematic representation of the computational model.



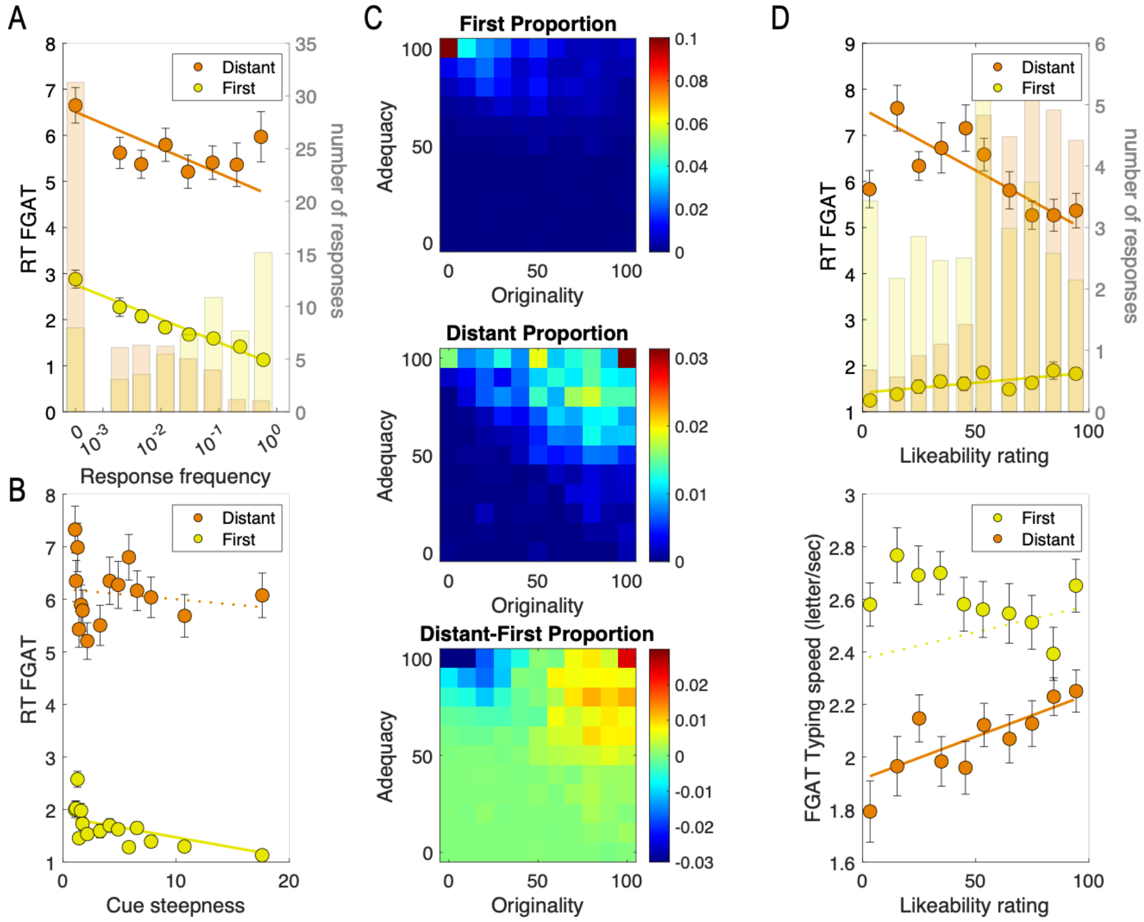
The model takes as input a cue, that “activates” a semantic memory network. Semantic search (exploration) is implemented as a biased random walk, in which node transition probability P is determined by the frequency of association F between the node i and its connected nodes j. The visited nodes (option 1 to n) are evaluated in terms of adequacy (A), originality (O) and the valuator assigns a likeability (L) to each of them, CES stands for Constant Elasticity of Substitution, see Results. A response is selected in function of the FGAT condition: in the *First* condition (F), the selection is based on adequacy and in the *Distant* condition (D), the selection is based on likeability. Equations results from the different model comparisons conducted in the study and are detailed in the manuscript. Text in black corresponds to our framework and hypotheses while text green corresponds to the results obtained in our study.

Figure 2: Experimental design



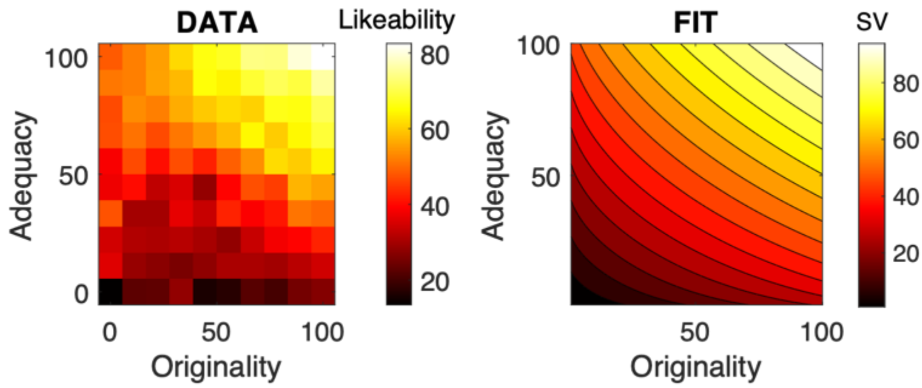
A. Chronological order of successive tasks. **B**. From top to bottom, successive screen shots of example trials are shown for the three types of tasks (left: FGAT task, middle: choice task, right: rating tasks). Every trial started with a fixation cross, followed by one cue word. In the **FGAT** task, when participant had a response in mind, they had to press the space bar and the word "Response?" popped out on the screen. The FGAT task had two conditions. Participants had to press a space for providing the first word that came to their mind in the *First* condition and an unusual, original but associated word in the *Distant* condition. In the **choice** task, two words were displayed on the screen below the cue. Participants had to choose the association they preferred using the arrow keys. As soon as a choice was made, another cue appeared on the screen and the next trial began. In the **rating** tasks, one word appeared on the screen below the cue. Then a scale appeared on the screen, noticing subjects that it was time for providing a response. In the likeability rating task, participants were asked to indicate how much they liked the association in the context of FGAT-distant. In the adequacy and originality rating tasks, each association was first rated on either adequacy and originality and then on the remaining dimension. Order was counterbalanced (see Methods for details).

Figure 3: Behavioral results of the FGAT task.



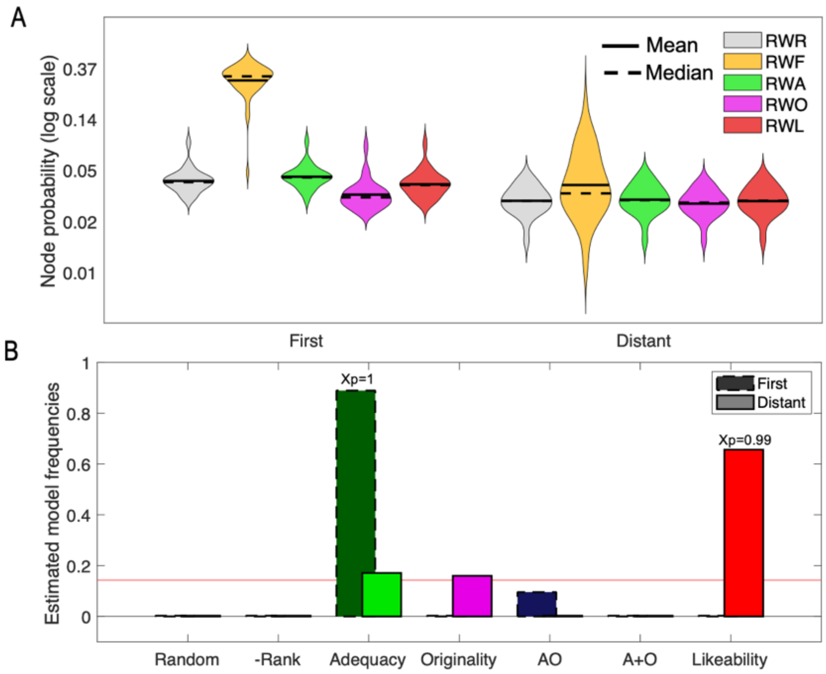
A. Correlation between response time (RT) in the FGAT task and the response frequency for the *First* (yellow) and *Distant* (orange) conditions. **B.** Correlation between response time (RT) in the FGAT task and the cue steepness for the *First* (yellow) and *Distant* (orange) conditions. **C**. Heatmaps of First (top), Distant (middle) and Distant-First (bottom) proportions of responses per bin of adequacy and originality ratings. **D**. Correlation between response time (top) and typing speed (bottom) in the FGAT task and likeability ratings of the FGAT responses for the *First* (yellow) and *Distant* (orange) conditions. In **A, B, D**, circles indicate binned data averaged across participants. Error bars are intersubject s.e.m. Solid lines correspond to the averaged linear regression fit across participants, significant at the group level (p<0.05). Dotted lines indicate that the regression fit is non-significant at the group level (p>0.05). In **A** and **D top**, transparent bars correspond to the average number of responses per bin of frequency (A) or likeability (D).

Figure 4. Behavioral results of the rating tasks: building the valuator module



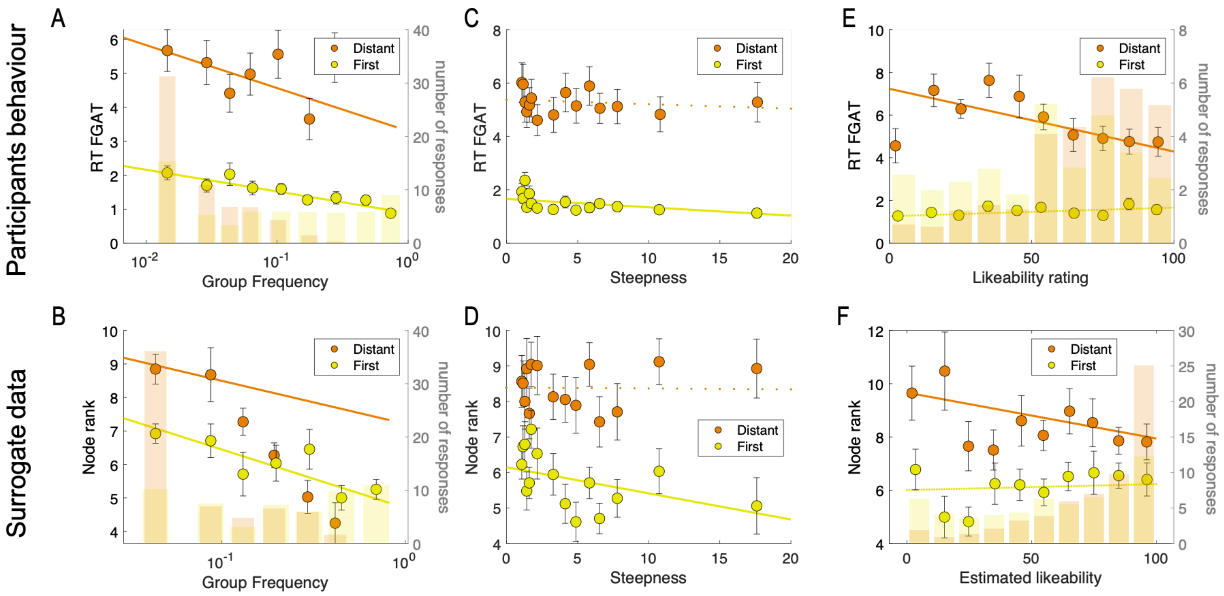
Average likeability ratings (left) and fit (right) are shown as functions of adequacy and originality ratings. Black to hot colors indicate low to high values of likeability ratings (left) or fitted subjective value (SV, right). The value function used to fit the ratings was the CES utility function.

Figure 5: Random walks predictions and selection model comparison.



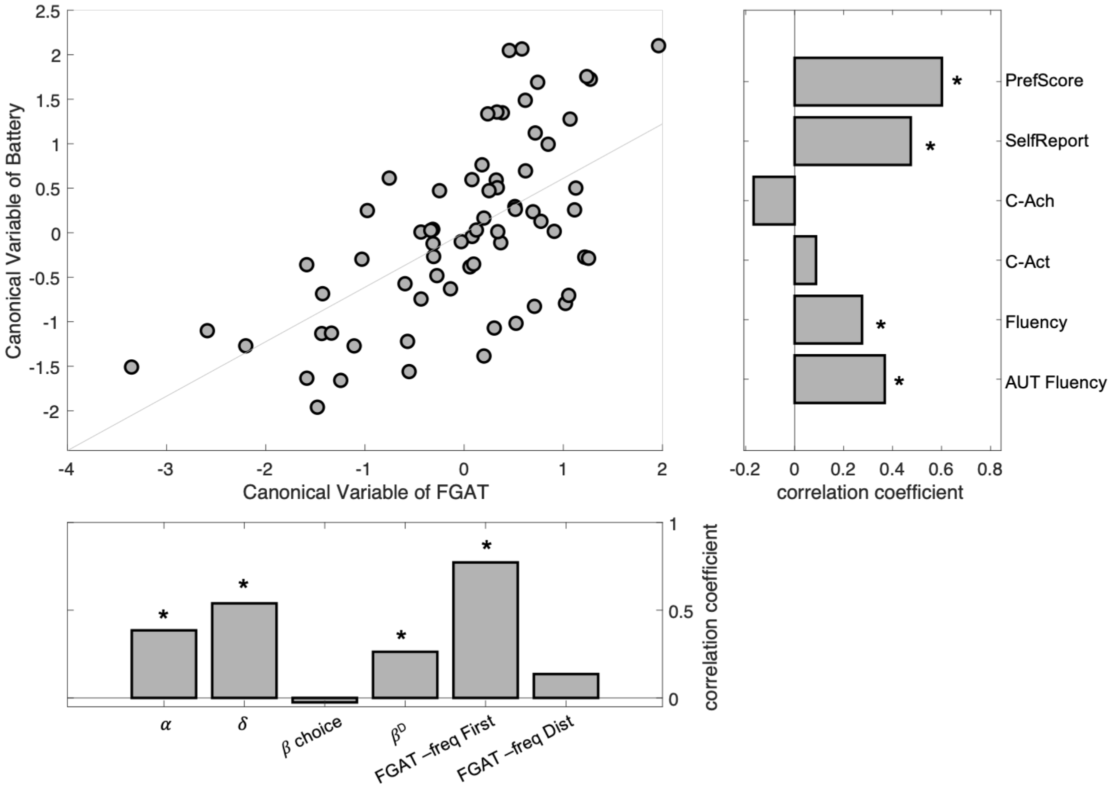
A. Violin plots of the probability of each random walk (RW) to reach the *First* and *Distant* participant responses in the semantic networks. RWR: random, RWF: frequency biased, RWA: adequacy-biased, RWO: originality biased, RWL: likeability biased. Violins represent the distribution of the averaged probabilities across trials for the subgroup of participants used to develop the model (n=46). **B.** Estimated model frequency of selection models for *First* (dark colors) and *Distant* (lighter colors) responses. Red line indicate chance level.

Figure 6. Response speed for the participants and surrogate data of the test group (n=23)



A, B. Correlation between response time RT (A) or node rank (B) in the FGAT task and the response frequency for the *First* (yellow) and *Distant* (orange) conditions. **C, D.** Correlation between response time RT (C) or node rank (D) in the FGAT task and the cue steepness for the *First* (yellow) and *Distant* (orange) conditions. **E, F**. Correlation between response time RT (E) or node rank (F) in the FGAT task and likeability ratings (E) or estimated likeability (F) of the FGAT responses for the *First* (yellow) and *Distant* (orange) conditions. Circles indicate binned data averaged across participants. Error bars are intersubject s.e.m. Solid lines corresponds to the averaged linear regression fit across participants, significant at the group level (p<0.05). Dotted lines indicate that the regression fit is non-significant at the group level (p>0.05). In **A, B, E** and **F**, transparent bars correspond to the average number of responses per bin of frequency (A, B) or likeability (E, D). Note that the surrogate data presented in the Figure correspond to one simulation (among 100) that is representative of the statistics obtained over all simulations and reported in the text.

Figure 7: Canonical correlation between the FGAT parameters/metrics and creativity tests belonging to a battery



Top left. Correlation between the first canonical variables of the battery of tests and of the FGAT parameters/metrics. Each dot represents one participant. Top right: correlation coefficient between each battery test and the canonical variable of Battery. Bottom left: correlation coefficient between each FGAT parameters/metrics and the canonical variable of FGAT. Stars indicate significance (p>0.05).