Reconceptualizing imitation in social tagging: a reflective search model of human web interaction

Paul Seitlinger
Knowledge Technologies Institute
Graz University of Technology
Graz, Austria
paul.seitlinger@tugraz.at

Tobias Ley
School of Digital Technologies
Tallinn University
Tallinn, Estonia
tley@tlu.ee

ABSTRACT

We analyze psychological dynamics of human-Web interaction exemplified by social tagging. Whereas previous models assumed tagging was driven by individual knowledge and social imitation, we introduce a reflective search framework that assumes user behavior (e.g., exploration and tagging of web resources) to arise from an iterative search of human memory shaped continuously by past and present learning episodes. We formalize this framework by means of a mathematical model of search of human memory which interrelates episodic and semantic memory processes. This allows us to simulate both temporal macro dynamics (stabilization of tag distribution) and underlying temporal micro dynamics (reflecting and tagging a resource). While the former are well covered by previous models, these models are not able to explain the latter. We claim that shifting away from imitation to reflective search holds great potential for understanding and designing human web interaction more generally, and to validate models of human memory in large-scale web environments.

CCS Concepts

•**Information systems** → *Web searching and information discovery;* •**Human-centered computing** → *Web-based interaction;* •**Applied computing** → *Psychology;*

Keywords

Social Tagging, Semantic Stabilization, Organism-Environment Dynamics, Search of human memory

1. INTRODUCTION

Social tagging on the Web allows studying social and cognitive processes on a large scale, such as memory, categorization and language use [7]. From a psychological research perspective, this is interesting as previous experimentation has been mainly lab-based, involving smaller samples of participants and creating controlled conditions to study these phenomena in depth. All these have to some extent limited the ecological validity and generalizability of results [11]. For example, when looking at memory processes in artificial individual learning trials in the lab, it may be overlooked that human memory is predominantly a tool used to interact with our social and material environment [2].

Making use of social tagging data for identifying regularities in how humans interact with the Web has been on the Web Science agenda for a number of years. Our reading of this literature is that mainly two directions have been pursued. The first has looked at collective stabilization and dynamics, and in particular at how semantic stabilization happens despite limited central control (e.g., [26],[9]). In this strand of research, different forms of imitation have played a major role, that is models of how users of the tagging system imitate other users' tags, and how agreement about the use of particular tags results from this (e.g., [8],[3],[9],[4],[28]).

The second line of research has been trying to establish regularities of how individuals learn over time (e.g., [5],[24]), and which tags they use in a particular moment to describe a resource [14]. This has been important to suggest tag recommender services that support users in the process of tagging (e.g., [14]). Imitation has also played a role in this second strand of research, e.g., in the semantic imitation model of Fu and colleagues (e.g., [6],[5]).

If we want to leverage the massive data available in social tagging systems for bringing forward psychological research, then it is important to critically review the theoretical basis on which modelling of social tagging data rests. When doing so, it seems to us that the focus on imitation in both strands of social tagging research is to some extent surprising. The focus on imitation seems to suggest that the main motivation of users of a social tagging system is that of imitating others. It seems as if people were constantly making decisions of whether they should reuse another person's tag or not. And in fact, current generative models of social tagging separate imitation from the use of prior (also called "background") knowledge that the user brings to the situation into two independent processes: I either draw on my existing knowledge, or I copy a tag from someone else. As we will show in section 2.2, this assumption has already been empirically challenged in several studies conducted in the tradition of the co-evolution model (e.g., [13]), and the semantic imitation model (e.g., [5]).

From a psychological point of view, this artificial separation has obstructed the view on social tagging as a process of searching and making sense of Web resources in which semantic and episodic memory together produce a particular

WebSci '16, May 22-25, 2016, Hannover, Germany

DOI: http://dx.doi.org/10.1145/2908131.2908157

set of tags in a concrete situation. While semantic memory represents more general sedimendations of cultural meaning making (e.g., whether Paris is the capital of France), episodic memory represents personal experiences (e.g., my visit to Paris last year and memories from this; e.g., [27]).

In this paper, we therefore suggest a model that we call "reflective search". We start from the assumption that instead of imitating others, the main motivation for someone to use a social tagging system is to support searching, collecting and organizing Web resources. By doing so, the model moves the focus away from imitation, and conceptualizes social tagging as a search process and its reflections in memory.

We will show that despite having no explicit imitation component, the model nevertheless produces social stabilization as a side effect because users are part of a shared meaning making system. At the same time, and because of its roots in memory search, the model also allows more fine grained predictions about the individual use of tags that other models have not been able to make. The model therefore unifies the distinct research traditions (on individual and collective processes) and integrates semantic and episodic memory processes into one single framework.

In the next chapter, we first review research in social tagging. We then propose the reflective search model and its mathematical formalization. This formalization allows us not only to validate the model with empirical data from a large-scale social tagging system, but also to construct a generative model of search and tagging. We subsequently present a multi-agent simulation study that is able to reproduce several well-known individual and collective phenomena and their relationship.

2. COLLECTIVE AND INDIVIDUAL PHENOMENA IN SOCIAL TAGGING

2.1 The role of imitation and knowledge in social tagging

The seminal work of Golder & Huberman [8] introduces a 2-factor, explanatory scheme of stable patterns in tag proportions and gives a subsequent discourse on self-organizing mechanisms in social tagging a strong and outlasting spin. The first factor they put forward is shared background knowledge that is assumed to cause different users to make similar tag choices for a given Web resource with or without being aware of others' behavior. Thus, from a cognitive viewpoint, [8] focus on a user's semantic-lexical memory, which is shaped by statistical structures of natural language and mirrors commonsense word associations [21]. As a second factor, their scheme invokes the notion of imitation as some kind of social motive that would drive users to seek for "social proof" and to regard popular tag recommendations as the "right" way to categorize the given resource.

The stochastic model that Golder & Huberman [8] suggest to formalize the quick development of stable tag proportions is restricted to the social factor of imitation. Implemented in form of a variant of Polya's urn model, it realizes a preferential attachment mechanism and successfully reproduces stable tag proportions typically observed in social tagging systems (e.g., Delicious).

One ambiguity of the 2-factor scheme is that [8] leave out a specification of how the two factors are interrelated both theoretically and formally. This may have well given rise to the aforementioned artificial separation of knowledge and imitation in how the field further developed. All subsequent works either continued putting a formal focus on imitation and neglecting the impact of knowledge, or they formally implemented the explanatory scheme without a theoretical reflection on the relationship between the two factors.

The works of [3] and [9] mainly attend to the factor of imitation and suggest refinements of the preferential attachment mechanism to better account for the characteristics of stabilized rank-frequency distributions of tags [3] and for the temporal dynamics in the stabilization of these distributions [9]. [3] adds a power-law of forgetting to simulate a user's preference for tags used frequently and recently within the system; [9] extends the mechanism by a sensitivity towards a tag's information value to simulate preferences for frequent tags referring to concepts, which are neither too specific nor too general in describing the resource.

Due to the inability of all these models to account for a particular pattern, namely, the non-linear growth rate of unique tags (e.g., [4], [18]), [4] were the first who proposed to combine imitation with the stabilizing factor of shared background knowledge, i.e., to implement the 2-factor scheme of [8]. In their proposed "epistemic dynamic model" including five free parameters, [4] use an imitation parameter to let a simulated user either draw on her/his background knowledge (by sampling a word from a topic-related Web corpus) with probability I or choose a word from the currently most popular tags with probability 1-I. In the sense of [16], this proposed 'either/or' logic assuming stochastic independence of knowledge and imitation exhibits high outcome but – as we believe – low process validity. Even in most recent works, e.g., [28], the 'Background plus imitation' model is still widely used to explain semantic stabilization. In summary, although this model accounts for the emergence of stable patterns in social tagging, its underlying logic does not square well with a series of empirical studies on individual learning and human-Web interactions in social information systems, as we will show next.

2.2 Coupling of individual search and collective memory in tagging

First, research of [13] and [15] reveals a coupling of individual knowledge and joint artifacts (e.g., wiki articles, tag distributions): Whether a user's interaction with the joint artifact results in a qualitative extension (e.g., by adding a semantically new tag) or merely quantitative change (e.g., by imitating existing tags) depends on the user's prior knowledge, in particular, on the congruity between the individual understanding (e.g., resource interpretation) and the semantics conveyed by the artifact (e.g., recommended tags). Similarly, the 'Semantic Imitation' model of Fu and colleagues that is based on empirical studies (e.g., [6]) and model-based analyses (e.g., [5]) implies strong dependence of imitation on the way users interpret Web resources: if people categorize an object differently, they are less likely to choose similar tags. In other words, the more people converge in their background knowledge, the stronger is the impact of imitation on the emergence of tag distributions.

Taken together, empirical evidence does not allow to treat imitation and knowledge as two stochastically independent factors, and if the goal of Webscience is not only to achieve high outcome but also high process validity, such evidence should motivate the accommodation of established models of social tagging.

From our perspective, the semantic imitation model [5] provides an appropriate starting point for the endeavor of refining dynamic models of social tagging. As a sophisticated model of sensemaking built upon a scientifically sound theory of human categorization [1], it models the formation and refinement of mental categories during a user's exploratory search of the information ecology within a social tagging system. By accounting for individual, episodic learning processes as well as for the categorization of a present Web resource, it also allows to anticipate whether the corresponding user complies with existing tags or whether she/he introduces a new tag to express an idiosyncratic resource categorization. That way, it establishes a link between individual-episodic learning and the emergence of macro structures, e.g., stabilized tag distributions. However, users simulated by the semantic imitation model have "no [prior] commonsense knowledge" ([5], p. 44) and thus, the model lacks the integration of episodic learning into background knowledge, i.e., of a pre-existing semantc-lexical memory that gets shaped during an exploratory information search and supports inter-individual sensemaking. For a model that should anticipate dynamics on a macro level, we assume that it is crucial to consider such prior semantic knowledge and how it interacts with episodic learning during exploratory search. E.g., depending on the environment to be observed and understood (e.g., Delicious vs. Pinterest), users can be more or less heterogeneous in terms of pre-existing semantic memory, and the extent of heterogeneity in semantic memory can be assumed to affect future dynamics and stabilization.

Based on this brief overview of dynamic models of social tagging, we identify two major gaps that we will describe next and address in this paper by introducing a unifying model of 'Reflective Search' to be validated in an empirical study and model-based analysis.

2.3 Contributions of a reflective search model

The first gap identified is that existing approaches put emphasis either on commonsense semantic memory structures ([8],[3],[9],[4]) or individual and episodic learning processes (e.g., [5]). Therefore, the first goal is to introduce a unifying 'reflective search' framework that helps to understand how prior semantic memory and episodic learning (during a user's previous Web navigation) interact when users reflect upon a Web resource and choose tags to index their reflections. That way, tagging becomes a verbalized epiphenomenon embedded naturally into a cognitive user-resource loop, and imitation becomes a correlate of inter-subjectivity that refers to an inter-individual state of reflective agreement with respect to a resource.

The second gap covered by our framework concerns the temporal micro dynamics that unfold as a user reflects upon and tags a resource. Previous models have targeted temporal macro dynamics distributed across consecutive bookmarks of different users. The resulting tag resource stream [4] has then been analyzed in terms of a stabilizing tag frequency distribution and quantified applying measures, such as Kullback-Leibler Divergence [9], Rank-Biased Overlap [28], the rate at which the number of unique tags increases [18] and tag proportions stabilize [8], or simply, calculating the probability of a bookmark including a new tag [25]. From all these studies, however, little has been learned about the temporal micro dynamics underlying a single tag assignment (TAS) of a given user. While the seminal work of [8] has already included a cognitive and empirical consideration of these dynamics, it has not been taken up in subsequent models. In particular, [8] reveal early tags in a TAS to be basic-level tags shared by many users, and tags at later TAS positions to be idiosyncratic. Though it's immediately obvious that micro and macro dynamics should be interlinked – as basic level tags should drive semantic stabilization – until now, no model has been developed that accounts for temporal micro dynamics.

Therefore, the goal of this paper is i) to introduce the reflective search framework and a corresponding, mathematical model of human memory search that is called Context Maintenance and Retrieval (CMR; e.g., [22],[19]) as a first concretization, and ii) to demonstrate its ability to close these two gaps by means of model-based simulations of empirical patterns extracted from a Delicious dataset. In the following, we first describe and second, test the model.

3. REFLECTIVE SEARCH MODEL

3.1 Model Overview

Figure 1 shows a scheme of our 'reflective search' framework that should help to clarify how it is applied to realize two main model requirements: i) modeling the way a user's personal history, i.e., Web navigation, shapes pre-existing, commonsense associations (semantic memory component) through learning experiences (episodic memory component), and ii) modeling how the user's evolved memory (including episodic and semantic components) supports the micro dynamics during reflecting and tagging a new Web resource.

Episodic learning during Web navigation. First, the figure shows two layers, which mutually influence each other. There is a manifest layer, which can be observed (e.g., in log files) and represents a user's sequence of collected resources (black dots). Please note that, within the notational system of CMR, internal stimuli (e.g., thoughts) and environmental stimuli (e.g., Web resources or tags) are called items and denoted as $\mathbf{f}$. Thus, to refer to a Web resource (e.g., article), we use the symbol $\mathbf{f}_i$, where the subscript i indicates the resource's position in the user's resource sequence. Additionally, the model includes a latent layer, which is unobservable and represents a sequence of context states (grey dots) within a user's memory. In the sense of [22], we use the notion of context to refer to a blend of memories, which is evoked by a Web resource (e.g., memories of similar articles) at position i and represented by the symbol $\mathbf{c}_i$. This retrieved context determines two kinds of search processes: i) an 'internal' memory search for further memories (i.e., internal items) that – in the present study – is regarded as a reflection upon a resource, and ii) an environmental Web search for a further resource (i.e., environmental item $\mathbf{f}_{i+1}$). In the figure, the internal search process is illustrated by the schematic helix spanned between the last black and grey dot. The specific mechanism underlying this reflective and iterative process that we also use to simulate tag choices will be explained in detail below. The environmental search process is indicated by the dashed arrows (e.g., from $\mathbf{c}_i$ to $\mathbf{f}_{i+1}$).

Human memory is a very plastic, neural structure, keeping traces of each episodic experience (e.g., [10]). Therefore, searching memory or the Web is a cumulative learning process, which always depends on past and current events, such

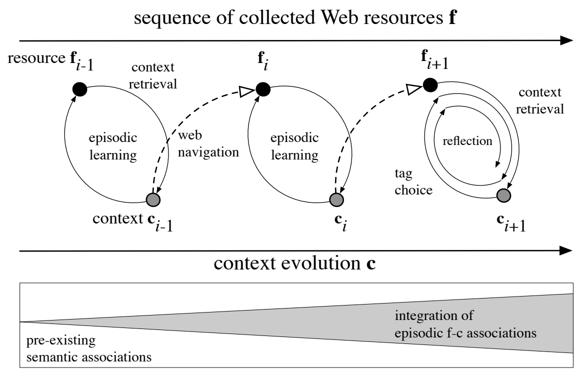


Figure 1: Reflective search scheme

as consecutively collected and reflected Web resources. In order to capture the learning processes during Web navigation, the model includes malleable associations storing item-to-context associations (to retrieve resource-relevant context) and context-to-item associations (to search for further context-relevant items). In this study, each episode (reflecting and tagging a Web resource) results in new episodic associations between the current resource ($\mathbf{f}_i$) and the current context ($\mathbf{c}_i$), which are integrated into former episodic associations and added to pre-existing, semantic associations. In Figure 1, the continuous integration of episodic learning experiences into pre-existing, semantic associations (first modeling requirement) is illustrated by the grey area, which increases in size along a user's Web navigation. Thus, intersubjectivity, i.e., the extent to which different users agree on their reflection upon a resource and are inclined to imitate each other, depends on the impact (relative strength) of pre-existing semantic knowledge and on users' convergence in previous Web navigation (personal learning episodes).

An essential construct in CMR is context evolution. The current context $\mathbf{c}_i$ is not only a blend of memories probed by the current resource $\mathbf{f}_i$ but it also blends with temporally weighted context states retrieved by previous resources (or internally retrieved items). Hence, it refers to an internally maintained context representation (activation pattern) in a user's memory that changes continuously in time (e.g., during Web navigation). For example, when I'm reading an article found on the Web, my thoughts and reflections upon it are not only affected by the memories evoked by that present article but also by memories and associations probed by previously encountered resources (e.g., other articles, videos, etc.). In terms of CMR, the evolved context state plays the role of an internal 'spotlight' that guides the search of memory by 'illuminating' context-relevant items, such as words that a user might use as search terms to navigate to a new Web resource or as tags to index her or his resource reflections. The latter micro-dynamics of reflecting upon a resource and using tags to externalize these reflections (second modeling requirement) is described next.

Micro dynamics in reflecting and tagging a Web resource. The schematic helix in Figure 1 (circulating between $\mathbf{f}_{i+1}$ and $\mathbf{c}_{i+1}$) illustrates the micro-dynamics that we assume to underlie a resource reflection accompanied by corresponding tag choices. In a first iteration t, a new resource (environmental item) addressing particular semantic categories (e.g., 'history' and 'humanities') probes item-to-context associations and retrieves new context (e.g., memories of articles dealing with similar categories) to be integrated into the already evolved spotlight, which then probes context-to-item associations to retrieve an internal item, such as a word semantically related to the spotlight. We assume the user to apply this retrieved word as a first tag. The retrieved word continues the resource reflection by probing new item-to-context associations in a subsequent search iteration t+1 and retrieving new context, which causes a slight spotlight shift, in turn guiding the search for a new word, i.e., tag. This iterative search proceeds until the spotlight stops shifting sufficiently to let the user become aware of new semantic aspects of the resource, i.e., until reflection comes to an end.

3.2 Model-based predictions and hypotheses

Based on this theoretical analysis of temporal micro dynamics within a single tag assignment (TAS), we derive a first prediction that does not follow from any of the existing models and helps to relate aspects of semantic stabilization (e.g., [28]) to the basic-level effects in TAS observed by [8].

In particular, *we predict that for any pair of users, the probability of choosing similar tags for a given resource decreases along the consecutive positions within a TAS.* In other words, users are more likely to agree on earlier than on later tags when describing the content of a resource. We make this prediction, because according to CMR, the very first search iteration of each user's resource reflection is cued by the same environmental item (i.e., Web resource), which retrieves comparatively similar context to be integrated into the internal spotlight, in turn guiding a relatively uniform search for an internal item to be translated into the first tag. In a later iteration t, however, the spotlight is already a blend of contextual states, which have been updated by the first environmental and the subsequent (t-1) internally retrieved items. As the probability of two users' iteration sequences including different items gets larger with an increasing sequence length, we can expect that the spotlights and hence, tag choices of later iterations (i.e., positions in TAS) diverge more strongly between users than tag choices of earlier iterations. Hence, we attribute the basic-level observation of [8] to the assumption that users' internal spotlights drift when reflecting upon a resource.

We will test this 'drifting spotlight assumption' by analyzing the stabilization of a tag resource stream separately for early and later positions of consecutive TAS. As already mentioned in section 2.3, stabilization of tag resource streams has been characterized in different ways. Among others, it is reflected by a decreasing probability of consecutive bookmarks' TAS including a new tag (e.g., [25],[5]). More specifically, if $p_{new}(r)$ denotes the probability of the rth bookmark's TAS including a new tag, then [25] and [5] have shown that $p_{new}(r)$ decreases along consecutive bookmarks r. If we further let t indicate the position of a tag within a TAS, *the first and so-called 'drifting spotlight' hypothesis H_{DS} is that – independent of r – $p_{new}(r,t)$ increases monotonically with an increasing t.* In other words, a tag added by a user is more likely being new if it appears later in the user's set of chosen tags. If H_{DS} holds, we can derive the further expectation that temporal dynamics on the micro level (iterative search of memory) are coupled with those on the macro level (emerging tag distribution) and thus, that stabilization is more strongly pronounced for tags in early than for tags in later TAS positions. In particular, *the second so-called 'micro macro coupling' hypothesis H_{MMC} is that the exponential decline (slope) of $p_{new}(r,t)$ along consecutive bookmarks r, which has already been observed by [25], decreases with increasing t.*

We will test and validate the model by observing i) whether the hypotheses comply with empirical data and ii) whether a multi-agent simulation, in which each agent behaves according to CMR assumptions, yields estimates of $p_{new}(r,t)$ for different r and t that fit empirical estimates well. In contrast to all prior work on tagging models (except for [5]), our model validation goes beyond a purely qualitative comparison of predicted and observed data and instead, evaluates the model's goodness of fit quantitatively.

Though other and more refined measures of stabilization have been applied (e.g., [9] or [28]), in this study we draw on the measure of $p_{new}(r,t)$ as it yields a pattern of frequencies that is most suitable for our parameter fitting technique and goodness of fit test (see section 4.1).

3.3 Generative mechanism

In the multi-agent simulation reported below, every simulation run involves a number of m agents each completing a first episodic learning and a second social tagging phase. In the first phase, a real user history (with a sequence of at least 20 Web resources) is sampled from a Delicious dataset [29] and used to train the given agent according to CMR model equations (see next section). Then, in the second phase, this trained agent is exposed to a set of new resources (and associated tags assigned by previously active agents) in order to reflect upon and tag each of them. That way, theoretical tag distributions emerge in the course of the simulated, social tagging phase, whose characteristics, e.g., inter-agent agreement in tag choices, or $p_{new}(r,t)$, can be compared with characteristics of empirical distributions.

The dataset, on which we draw, includes bookmarks of Wikipedia articles that we regard as environmental items, each characterized by one or several predefined Wikipedia top-level categories that have been assigned by the Wikipedia community. In total, there are 25 categories, and we apply them to represent a given article as a feature vector $\mathbf{f}$, which, according to CMR notations, is a standard basis vector "of unit length with a single non-zero element" ([19] p. 339). To comply with this form of notation, we let each element j in $\mathbf{f}$ correspond to one of a total of J=891 unique category combinations, where the vector's only non-zero element indicates the combination of categories assigned to the article.

Formal structures and dynamics.

The two layers illustrated in Figure 1 are represented by the feature layer F and the context layer C, including J=891 elements each. In the first episodic learning phase, an agent passes through the sequence of articles (environmental items) collected by the user assigned to the agent. The states of the two layers F and C change along this sequence and are represented at a particular position i by the two vectors $\mathbf{f}_i$ and $\mathbf{c}_i$, respectively. In the learning phase, $\mathbf{f}_i$ is given by the present article (i.e., its category combination j). The two vectors $\mathbf{f}_i$ and $\mathbf{c}_i$ mutually influence each other, and this item-context communication is mediated by the two matrices M^{FC} and M^{CF}, storing strengths of associations from item to context (M^{FC}) and from context to item elements (M^{CF}). Both matrices consist of a pre-existing, semantic component (M^{FC}_{pre}, M^{CF}_{pre}) and an episodic component (M^{FC}_{epi}, M^{CF}_{epi}). To simplify matters, M^{FC}_{pre} is initialized to an identity matrix, and the semantic associations are built into M^{CF}_{pre} (e.g., [19],[22]). Every element (a,b) in M^{CF}_{pre}, i.e., the strength of association between the category combinations a and b, is determined by calculating the *Jaccard Index* $JI(a,b)$ scaled by the free parameter s. E.g., if a includes 'history' and 'humanities' and b includes 'history' and 'language', $JI(a,b)$ is 1/3. The episodic components are initialized to 0, and the aim of the episodic learning phase is to shape them by user-specific associations as well as to give rise to a user-specific context vector (spotlight) $\mathbf{c}$. This learning process is formalized by CMR equations (1)-(5b) validated through several psychological experiments.

To simplify matters, we have implemented CMR's core assumptions and left out processes that are of minor relevance for our research questions. In particular, we have not considered the primacy effect (memory advantage for early learning episodes), resulting in a simplification of Equation 5b. As our research questions do not address inter-response times of consecutive tags in a TAS, we have also abstracted from detailed assumptions on selecting tags for output (Equation 8). Beyond that, Equation 9 and 10 are not included in CMR and are applied in this work to model a tagging-specific process (i.e., the choice probability of a tag as an interplay of reflection and previous users' tag choices).

Episodic learning phase.

Context evolution. Consider an agent encountering an article at position i (in the corresponding user's bookmark sequence), which evokes a patterned activation in the feature layer F, i.e, $\mathbf{f}_i$, and provides contextual input $\mathbf{c}_i^{\mathbf{IN}}$ to layer C given by

$$\mathbf{c}_i^{\mathbf{IN}} = \frac{M^{FC}\mathbf{f}_i}{\left\|M^{FC}\mathbf{f}_i\right\|} \tag{1}$$

that is integrated into the context state $\mathbf{c}_i$ through the context evolution equation:

$$\mathbf{c}_i = \rho_i \mathbf{c}_{i-1} + \beta_E \mathbf{c}_i^{\mathbf{IN}} \tag{2}$$

The drift parameter β_E determines the rate (during the episodic learning phase) at which newly retrieved context (i.e., $\mathbf{c}_i^{\mathbf{IN}}$) is integrated into the context state $\mathbf{c}_i$. ρ_i is calculated at each position i to weight the previously updated context state and to ensure that $\left\|\mathbf{c}_i\right\| = 1$. It is given by

$$\rho_i = \sqrt{1 + \beta_E^2[(\mathbf{c}_{i-1} \cdot \mathbf{c}_i^{\mathbf{IN}})^2 - 1]} - \beta_E(\mathbf{c}_{i-1} \cdot \mathbf{c}_i^{\mathbf{IN}}) \tag{3}$$

Forming episodic associations. Hebbian outer-product learning is applied to form new episodic associations between $\mathbf{f}_i$ and $\mathbf{c}_i$ [22] given by

$$\Delta M^{FC}_{epi} = \mathbf{c}_i \mathbf{f}_i^{\mathrm{T}} \tag{4a}$$

$$\Delta M^{CF}_{epi} = \mathbf{f}_i \mathbf{c}_i^{\mathrm{T}} \tag{4b}$$

where T indicates the transpose of $\mathbf{f}_i$ and $\mathbf{c}_i$, respectively.

To maintain this learning effect, the episodic associations from item to context (equation 4a) and from context to item elements (equation 4b) are integrated into M^{FC} and M^{CF}, respectively, according to the following weighted sums

$$M^{FC} = (1 - \gamma_{FC})M^{FC}_{pre} + \gamma_{FC}M^{FC}_{epi} \tag{5a}$$

$$M^{CF} = (1 - \gamma_{CF})(D + sM^{CF}_{pre}) + \gamma_{CF}M^{CF}_{epi} \tag{5b}$$

where γ_{FC} and γ_{CF} are free parameters controlling the relative strengths of newly learned associations. D is an identity matrix introduced to ensure that the on-diagonal associations are not affected by the semantic scale factor s.

Then, the next article (i.e., $\mathbf{f}_{i+1}$) is passed on to the agent to repeat these processes, i.e., equations (1)-(5b). Once it has processed all the user's articles, the agent has integrated user-specific episodic associations into pre-existing associations and has formed a unique context state. Next, we show how to make use of the evolved structures and states to simulate realistic tag assignments.

Social tagging phase.

Iterative search of memory: Reflecting upon an article. In this phase, each of the m trained agents assigns tags to the same set of n_a new articles. The consecutive assignment of tags to a given article is realized as an iterative search of memory, which is triggered by a newly presented article and involves several iterations t each yielding a tag the agent assigns to the article. In the first iteration ($t = 1$), the article evokes a pattern $\mathbf{f}_t$ in F and provides input to C via M^{FC} given by equation (1), where in this phase we apply the running index t instead of i.

Again, the newly retrieved context updates the context state according to the context evolution equation (2), where this time (i.e., during the tagging phase), the drift parameter depends on the number of iteration t:

$$\beta_t = \begin{cases} \beta_E & \text{if } t = 1 \\ \beta_I & \text{if } t > 1 \end{cases} \tag{6}$$

If $t = 1$, $\mathbf{f}_t$ represents an environmental item (article) and thus, the rate of context integration is controlled by the parameter β_E already applied in the episodic phase. However, in subsequent iterations ($t > 1$), $\mathbf{f}_t$ represents an internal item (see below) and thus, a different drift parameter, i.e., β_I, is applied to control for the rate of context integration.

Once the context state has been updated, $\mathbf{c}_t$ plays the role of a spotlight and guides the search for a new item stored in an agent's memory. In particular, $\mathbf{c}_t$ evokes an activation pattern $\mathbf{a}$ on the feature layer F given by

$$\mathbf{a} = M^{CF}\mathbf{c}_t \tag{7}$$

where $\mathbf{a}$ includes an activation value for each of the $J =$ 891 elements j in F, i.e., category combination. Following [20], the probability $P(j)$ of retrieving j is defined as

$$P(j) = \mathbf{a}_j^{\tau} / \sum_k^J \mathbf{a}_k^{\tau} \tag{8}$$

where τ is a free parameter controlling the sensitivity in $P(j)$ to activation differences. The category combination retrieved is denoted as j_t and becomes the internal item that cues the next iteration $t + 1$ that proceeds along equations (1), (2) and (6)-(8).

Indexing each search iteration t: Tagging an article. We assume that in each iteration, the retrieved (internal) item becomes manifest (e.g., visible for others), if the agent indexes the item's category combination j_t in form of a tag w. To model this indexing process, we extend CMR by a simple mechanism and let every agent have access to a lexicon L, which is the set of all tags w that have been generated by the m users in the Delicious dataset. We assume that the probability $P(w)$ of choosing the tag w depends on its semantic utility u_w affected by two interacting variables: the agent's reflection on the article $\mathbf{f}_1$ and the behavior of former agents that have already assigned tags to $\mathbf{f}_1$. Note that in the tagging phase, $\mathbf{f}_1$ refers to the present article as the subscript does not indicate the article's position but the current iteration of memory search. Referring to the first variable, in each iteration t, every tag takes on a particular semantic strength $p(w|j_t)$ that is approximated by the number of times w co-occurs with j_t divided by the total occurrence number of w in the entire Delicious dataset.

The second variable (behavior of previous agents) has an amplifying effect on the semantic strength (the first variable). We assume that the utility of a semantically appropriate tag w further increases, if it is presented to the agent via a tag recommendation mechanism. In Delicious, the probability of a tag w being recommended depends on its popularity, and in our model, we approximate it by $p(w|\mathbf{f}_1)$ that represents the relative frequency by which w has already been assigned to $\mathbf{f}_1$ by previous agents. Therefore, the semantic utility u_w is given by

$$u_w = p(w|j_t)[1 + p(w|\mathbf{f}_1)]^{\phi} \tag{9}$$

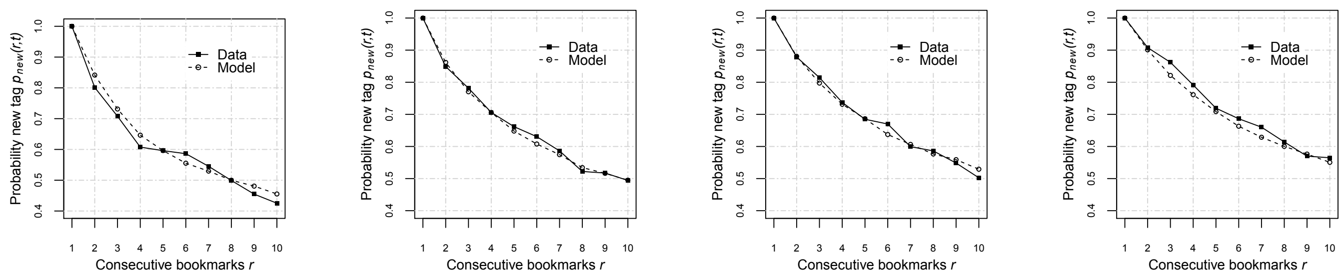
where ϕ is a free parameter controlling the extent by which a tag's popularity amplifies its semantic utility. Finally, the probability $P(w)$ of choosing tag w results from

$$P(w) = u_w / \sum_k^N u_k \tag{10}$$

Taken together, equations (9) and (10) represent the assumption that a tag's semantic utility and consequently, choice probability, is greater than 0, only if it complies with the agent's interpretation of the article (i.e., if $p(w|j_t) > 0$). Furthermore, if the tag is semantically appropriate and – due to inter-subjectivity – has also been assigned to the article by former agents (i.e., $p(w|\mathbf{f}_1) > 0$), the tag's utility and choice probability get larger.

4. MULTI-AGENT SIMULATION STUDY

A multi-agent simulation is a way to validate complex socio-cognitive mechanisms that are difficult to observe due to the underlying dynamical inter-relations of several variables. Typically, an empirical pattern is first identified (in our case, this is for one the micro level dynamics, the second the emerging tag distribution) and then, the generative mechanism is used to try to simulate the pattern. Hence, a multi-agent simulation not only allows for validating the outcome, but also allows for process validation by comparing the empirical with the simulated pattern. In the multi-agent simulation study described next, each of the m agents be-



(a) $t = 1$ (b) $t = 2$ (c) $t = 3$ (d) $t = 4$

Figure 2: Decline of $p_{new}(r,t)$ – the probability of a tag being new at a particular bookmark r and a particular position t within a tag assignment TAS – along consecutive bookmarks and for varying TAS positions

haves according to the generative mechanism, i.e., passes through an episodic learning phase (equations 1-5b) and subsequently, participates in a social tagging phase. That way, we can test our hypotheses more stringently and in two steps: first, we test them empirically by analyzing Delicious data and second, compare the empirical data with results of the simulation by performing a goodness-of-fit test.

In the tagging phase, a tag assignment (TAS) emerges as an agent reflects upon a given article along an iterative search of memory (equations 1, 2, and 6 - 8), where each iteration t results in a tag choice (equations 9 and 10). In line with the notations introduced in section 3.2, t indicates a tag's position within a TAS. In the multi-agent simulation, each article is tagged consecutively by all m agents, where r indicates the agent's position within this agent sequence. Thus, r corresponds to the rth bookmark of a resource within the Delicious dataset, and $p_{new}(r,t)$ represents the probability that a tag, which is chosen by the rth agent in iteration t, is new.

In section 3.2, we have derived the *'drifting spotlight'* hypothesis H_{DS} that estimates of $p_{new}(r,t)$ should increase monotonically along the TAS positions, i.e., by increasing t. The rational behind H_{DS} is that if t increases, the difference in context states between any pair of agents (or users) gets larger because during reflection, their spotlights drift and are increasingly less determined by a shared-environmental but more determined by eventually individual-internal items. As we assume that these temporal micro dynamics give rise to temporal macro dynamics (i.e., stabilization), the second *'micro macro coupling'* hypothesis H_{MMC} predicts that the slope of the exponential decline of $p_{new}(r,t)$ along consecutive r is larger for early than for later TAS positions t.

The questions that arise now are: i) can we actually observe an empirical pattern within the Delicious dataset that harmonizes with the hypotheses and ii) can we simulate a distribution of $p_{new}(r,t)$ for different r and t that does not diverge significantly from the empirical distribution?

4.1 Parameter fitting technique

Before describing the genetic algorithm by which we search the model's parameter space to find an optimal set of parameters, we first explain how we obtain the simulated and empirical frequencies underlying estimates of $p_{new}(r,t)$.

Simulated frequencies. In each simulation run, a computationally feasible number of $m = 10$ agents are involved and thus, for every new article to be tagged in the social tagging phase we get a sequence of $r = 1, ..., m$ bookmarks (and associated TAS), where we set the number of new articles to $n_a = 30$. Furthermore, we let each agent assign a number of $n_t = 4$ tags to every article, and thus, get a sequence of $t = 1, ..., n_t$ positions within a TAS. For the first agent tagging a given article, $p_{new}(1,t) = 1$. Therefore, the result of each simulation run is a 4(positions) $\times$ 9(bookmarks) contingency table, where each cell includes a count for the number of resources to which new tags have been assigned (at a given r and t). Then, $p_{new}(\mathrm{TAS}_{r,t})$ is estimated by dividing the count of cell (t,r) by $n_a = 30$. After obtaining the best-fitting parameter estimates (see *Genetic algorithm* below), we conduct 500 simulation runs and obtain the final estimates of $p_{new}(\mathrm{TAS}_{r,t})$ by determing the average 4×9 contingency table.

Empirical frequencies. We have extracted a number of $n_a = 1.046$ articles, for which there are at least $m = 10$ bookmarks each described by at least $n_t = 4$ tags. That way, we get a corresponding empirical 4×9 contingency table, in which each cell (t,r) includes the count for the number of resources to which new tags have been assigned at the corresponding TAS position t and bookmark number r. Again, we estimate the empirical $p_{new}(\mathrm{TAS}_{r,t})$ by dividing the count of cell (t,r) by $n_a = 1.046$.

Genetic algorithm. Following the CMR parameter fitting technique, we apply a genetic algorithm to find the best-fitting parameters. Here we draw on the R package GA [23]. This algorithm aims to minimize fitness values for populations of parameter sets, which we define as the sum of squared errors between model (simulated estimates of $p_{new}(r,t)$) and data (empirical estimates of $p_{new}(r,t)$) weighted by the SE of the data (see also [22],[19]). GA starts by generating an initial population of strings, which are randomaly generated parameter sets. Here, we set the population size to 500. Then, fitness evaluation takes place and the model-data divergence of each string is determined to select the fittest – in this study, top 5% – strings. In a subsequent phase of exploration, processes of mutation (randomly alterning the values of selected strings) and crossover (combining values of two selected 'parent' strings) allow to

Table 1: Summary of free parameters

	Description	Range	Estimates
β_E	*Context integration rate, environmental item*	0-1	0.913
β_I	*Context integration rate, internal item*	0-1	0.213
s	*Semantic scale factor*	0-5	3.422
γ_{FC}	*Relative strength of* M_{epi}^{FC}	0-1	0.356
γ_{CF}	*Relative strength of* M_{epi}^{CF}	0-1	0.244
ϕ	*Amplification factor*	0-10	8.773
τ	*Sensitivity parameter*	0-5	4.075

generate a further 'generation' of strings. These steps (population generation, evaluation, exploration) are repeated until a particular criterion of convergence is reached, where – in this study – "GA stops if a number of [200] generations has passed without improvement" [23].

The best-fitting parameter set yielded by GA is shown in Table 1 to be discussed in section 4.2. For validating the model, we have used this set of model parameters to conduct 500 simulation runs and to obatin the simulated contingency table (see above) underlying the estimates of $p_{new}(\text{TAS}_{r,t})$. To evaluate the model quantitatively, we have first multiplied each simulated $p_{new}(r,t)$ by the total number of empirical observations (i.e., 1.046) and then compared these simulated frequencies with the corresponding empirical frequencies using a χ^2 goodness-of-fit test. Given 36 (=4*9) datapoints and 7 free parameters, the critical goodness-of-fit statistic is $\chi^2_{crit}(29) = 42.56$, which should not be exceeded by our model if it fits data well.

4.2 Results and Discussion

Testing model predictions with behavioral data.

The 'drifting spotlight' hypothesis H_{DS} assumes $p_{new}(r,t)$ to increase monotonically with an increasing t (independent of r). Remember that $p_{new}(r,t)$ is the estimated probability that a tag, which occurs in the rth bookmark at iteration (or TAS position) t, is new. Figure 2 presents the results: The four diagrams show the decline of $p_{new}(r,t)$ along the first $r = 1, ..., 10$ bookmarks separately for the $t = 1, ..., 4$ TAS positions, respectively. We first describe the solid curves, which represent the empirical estimates. We can easily recognize that – independent of r – the average probability of introducing a new tag $\bar{p}_{new}$ appears to increase from left to right, i.e., from the first to the fourth TAS position t. The four corresponding means can be found in Table 2 (see columns 'Data' and '$\bar{p}_{new}$') clearly revealing a monotonous increase of $\bar{p}_{new}$ along the four positions. In line with H_{DS} we therefore conclude that users are more likely to agree on early than on later tags (of a TAS) when reflecting and tagging the content of a resource.

The second hypothesis H_{MMC} assumes that this positive relationship between inter-user agreement and TAS position is coupled with temporal dynamics on the macro level (tag resource stream) and gives rise to a more strongly pronounced stabilization at early than at later TAS postions. The solid curves plotted in the diagrams of Figure 2 square well with H_{MMC} as their slopes λ seem to decrease from $t = 1$ (Figure 2, diagram a) to $t = 4$ (Figure 2, diagram d). Drawing on [18] who have found that the rate at which

Table 2: Descriptive characteristics ($\bar{p}_{new}$ and λ) of empirical and simulated data points

	Data		Model	
	$\bar{p}_{new}$	λ	$\bar{p}_{new}$	λ
t=1	.580	.093	.584	.089
t=2	.639	.077	.633	.078
t=3	.669	.069	.665	.069
t=4	.708	.060	.700	.064

the number of unique tags increases follows an exponential function, we have obtained estimates of λ by fitting the theoretical function $p(r,t) = e^{-\lambda * r}$ to each of the four positions t. The amount of variance explained by this theoretical exponential is 95%, 96%, 97%, and 97%, for $t = 1$, $t = 2$, $t = 3$, and $t = 4$, respectively. Table 2 provides these estimates of λ (see columns 'Data' and 'λ'), and discloses that the slope indeed gets smaller as t gets larger. We therefore conclude in accordance with H_{MMC} that it is the early tags in a TAS which drive temporal macro dynamics, i.e., stabilization.

These analyses of behavioral data extracted from Delicious lend strong support to our two hypotheses H_{DS} and H_{MMC}. As the hypotheses' predictions follow naturally from the proposed reflective search model and seem to harmonize well with empirical patterns, we deem these results a first step towards process validation. We further validated the model with the generative mechanism embedded into the simulation study, whose results are reported next.

Simulating behavioral data for process validation.

First, further qualitative evidence for the model's ability to account for users' tagging behavior is provided by Figure 2. The dashed curves in each of the four diagrams represent the result of the simulation, i.e., the estimates of $p_{new}(r,t)$ that we obtain when averaging across the 500 simulation runs each based on the best-fitting parameter set shown in Table 1. Despite some deviations from the data points, the picture of a model comes to the fore that generates the two predicted (and empirically observed) characteristics of the data well: By approximating the data well, the simulated curves show that i) the average probability of introducing a new tag $\bar{p}_{new}$ increases along the four TAS positions t (hypothesis H_{DS}), and ii) that the slope decreases from $t = 1$ to $t = 4$ (hypothesis H_{MMC}). The summarizing descriptive values ($\bar{p}_{new}$ and slope λ for each of the four TAS positions) are shown in Table 2 (see column 'Model': $\bar{p}_{new}$ and λ, respectively) and further suggest only slight deviations between data and model. As for the empirical curves, we have derived estimates of λ by fitting the theoretical exponential to the simulated curve at each of the four TAS positions (explaining 94%, 96%, 97%, and 97% of variance for $t = 1$, $t = 2$, $t = 3$, and $t = 4$, respectively).

Empirical test of model fit.

Second, an evaluation of the goodness-of-fit using a χ^2 statistic provides evidence for the model's process validity: comparing the empirical and simulated estimates of $p_{new}(r,t)$ yields a value of $\chi^2(29) = 13.74$ that is far beyond the critical value of $\chi^2_{crit} = 42.56$ (see section 4.1) and allows to keep the nullhypothesis that the simulated and empirical curves depicted in Figure 2 do not differ significantly.

Third, having gained qualitative and quantitative evidence for the model's process and outcome validity, we can now describe and discuss the parameters and their estimates shown in Table 1 yielding theoretically plausible values. The first two parameters β_E and β_I show that the rate by which newly retrieved context is integrated into the internal spotlight (used to probe new context-to-item associations) is larger for context elements activated by an article (β_E) than for context elements activated by an internally stored item (e.g., tag; β_I). This difference seems plausible as a user should be more likely of shifting her or his spotlight after having read a new article than after having reflected upon the very same article. The comparatively large estimate of the third (semantic scale) factor s reveals that pre-existing semantic (commonsense) associations have a relatively strong influence on the learning process and the subsequent reflection and indexing of new articles. This result harmonizes well with previous studies on models of social tagging (e.g., [4]) suggesting that we have to consider users' background knowledge in order to model a non-linear increase of the number of unique tags, or, as in our case, the decline of $p_{new}(r,t)$. The moderate estimates of γ_{FC} and γ_{CF} indicate that – in line with [5] – knowledge being stored in M_{FC} and M_{CF}, respectively, gets shaped by episodic learning experiences (captured by M_{epi}^{FC} and M_{epi}^{CF}), where the relative strength of this episodic influence seems to be smaller than that of pre-existing associations (which equals 1-γ_{FC} and 1-γ_{CF}) between elements of the item and context layer (and vice versa). This could explain states of inter-subjectivity (inter-user agreement in reflecting on Web resources), which – according to our model – becomes manifest in a verbal imitation behavior, despite of episodic learning processes in the course of individual Web navigation. The high estimate of the amplification factor ϕ reveals a strong impact of others' tagging behavior (e.g., displayed by tag recommendation interfaces) on individual tag choices *but only if* this behavior complies with individual reflections upon the present article (see equation 9). This appears to be a natural process during verbalization: If a user is exposed to a recommended tag (e.g., via a recommendation interface) that matches a thought the user would like to express, then she or he should be very inclined to adopt (or imitate) that semantically 'matching' tag. Finally, the estimate of the sensitivity parameter τ indicates that we have to increase the signal-to-noise ratio in the activation of memory items, which is evoked by the spotlight (context state; equation 7), in order to derive a realistic activation-based probability of item retrieval (equation 8).

5. CONCLUSION

We have gained evidence speaking in favor of the process and outcome validity of the proposed reflective search model. First, based on the drifting spotlight assumption, we have successfully predicted a pattern of results that can't be derived from existing models of social tagging (H_{DS}) and contributes substantially to our understanding of stabilization (H_{MMC}). Second, the reflective search model allows simulating frequency distributions that closely fit empirical data, and third, it yields plausible estimates of parameters that help to interpret the interplay of internal processes (e.g., context update and episodic learning) and environmental cues (e.g., Web resources and recommended tags).

One aim of this work is to shed fresh light on a dominant conception of the notion of imitation within the Web-science discourse around socio-cognitive phenomena, such as social tagging. According to our proposed reflective search model, imitation is not a separate process detached from individual, associative memory structures that can be modeled as a stochastically independent factor (as e.g. proposed by the 'Background-Plus-Imitation' model [4],[28]). Instead, we deem it an integral process of an inter-subjective reflection upon content, in which those tags will be adopted and become popular that help index and verbalize inter-individually similar reflections. Evidence for the basic assumption that the choice (and imitation) of tags is simply an epiphenomenon of an iterative search of memory (that we assume to underlie reflection) comes from Delicious data, which comply with our first model prediction (the 'drifting spotlight' hypothesis H_{DS}) and with the model's ability to simulate patterns that fit empirical data well.

The exclusion of imitation as an independent factor is in line with works of e.g. [13] and [5] that suggest dependence (i.e., co-evolution and coupling) of (social) imitation and (individual) knowledge, where we believe that our approach refines these works by important aspects: In contrast to [5], our model includes structures and mechanisms by which episodic learning is integrated into prior background knowledge and thus, helps to understand how states of inter-subjectivity can come into being despite of diverging learning experiences. In contrast to [13], we endeavor to concretize and formalize a coupling of internal and environmental phenomena (e.g., memory and Web resources) by means of a contemporary model of search of memory (i.e., CMR [22],[19]) in order to reduce the gap between theoretical considerations and empirical observations.

We also see some potential to irritate conceptions of how the individual is coupled with the so-called social system and of how temporal micro and macro dynamics interact. A little in the manner of [17], we assume that the tempting but somehow artificial differentiation between 'the individual' and 'the social' becomes obsolete as soon as we stop conceiving and modeling the individual as a simple element within a complex whole. And indeed, our empirical and simulated results that provide evidence for the 'micro macro coupling' hypothesis H_{MMC} suggest that stabilization (as an emerging artifact of the 'whole') falls into place just by modeling the individual as a reflected being and by letting these beings interact.

We would like to stress that the reflective search scheme of Figure 1 is not limited to social tagging but frames our general way of thinking about human-Web interactions. Our main argument is that interacting on the Web shapes memory and the internal spotlight (context representation), and these cumulative learning processes drive future behavior, which can become manifest in many different ways, e.g., in choosing tags, in generating new search terms to navigate to the next page, in conducting an exploratory search for the purpose of information-based ideation [12], in browsing through and selecting recommendations, etc.

In future work we will therefore apply the reflective search scheme to model observable and latent correlates of Web navigation (e.g., search paths and accompanied internal context evolution), but also to design intelligent and creatively stimulating recommendation mechanisms. For instance, the model would allow to detect lengthy monotonous search and consumption periods (without substantial spotlight shifts), to which the mechanism could react by providing novel and context-disrupting stimuli that could help the user to escape a state of mental fixation [12].

6. ACKNOWLEDGMENTS

This work is supported by the Austrian Science Fund (FWF): P 25593-G22, P 27709-G22, and the EU funded project Learning Layers (Grant Agreement 318209).

7. REFERENCES

[1] J. R. Anderson. The adaptive nature of human categorization. *Psychological Review*, 98(3):409, 1991.

[2] L. Barsalou. Situated simulation in the human conceptual system. *Language and cognitive processes*, 18(5-6):513–562, 2003.

[3] C. Cattuto, V. Loreto, and L. Pietronero. Semiotic dynamics and collaborative tagging. *Proceedings of the National Academy of Sciences*, 104(5):1461–1464, 2007.

[4] K. Dellschaft and S. Staab. An epistemic dynamic model for tagging systems. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, pages 71–80. ACM, 2008.

[5] W.-T. Fu and W. Dong. Collaborative indexing and knowledge exploration: A social learning model. *IEEE Intelligent Systems*, 27(1):39–46, 2012.

[6] W.-T. Fu, T. Kannampallil, R. Kang, and J. He. Semantic imitation in social tagging. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 17(3):12, 2010.

[7] R. J. Glushko, P. P. Maglio, T. Matlock, and L. W. Barsalou. Categorization in the wild. *Trends in cognitive sciences*, 12(4):129–135, 2008.

[8] S. A. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *Journal of information science*, 32(2):198–208, 2006.

[9] H. Halpin, V. Robu, and H. Shepherd. The complex dynamics of collaborative tagging. In *Proceedings of the 16th international conference on World Wide Web*, pages 211–220. ACM, 2007.

[10] D. L. Hintzman. Minerva 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, 16(2):96–101, 1984.

[11] D. L. Hintzman. Research strategy in the study of memory: Fads, fallacies, and the search for the âĂIJcoordinates of truthâĂİ. *Perspectives on Psychological Science*, 6(3):253–271, 2011.

[12] A. Kerne, A. M. Webb, S. M. Smith, R. Linder, N. Lupfer, Y. Qu, J. Moeller, and S. Damaraju. Using metrics of curation to evaluate information-based ideation. *ACM Transactions on Computer-Human Interaction (ToCHI)*, 21(3):14, 2014.

[13] J. Kimmerle, J. Moskaliuk, and U. Cress. Individual learning and collaborative knowledge building with shared digital artifacts. *International Journal of Behavioral, Cognitive and Psychological Sciences*, 1(1):25–32, 2009.

[14] D. Kowald, P. Seitlinger, C. Trattner, and T. Ley. Long time no see: The probability of reusing tags as a function of frequency and recency. In *Proc. of WWW'14*, New York, NY, USA, 2014. ACM.

[15] B. Kump, J. Moskaliuk, S. Dennerlein, and T. Ley. Tracing knowledge co-evolution in a realistic course setting: A wiki-based field experiment. *Computers & Education*, 69:60–70, 2013.

[16] D. Larsen-Freeman and L. Cameron. Research methodology on language development from a complex systems perspective. *The Modern Language Journal*, pages 200–213, 2008.

[17] B. Latour. Network theory| networks, societies, spheres: Reflections of an actor-network theorist. *International Journal of Communication*, 5:15, 2011.

[18] T. Ley and P. Seitlinger. Dynamics of human categorization in a collaborative tagging system: How social processes of semantic stabilization shape individual sensemaking. *Computers in human behavior*, 51:140–151, 2015.

[19] L. J. Lohnas, S. M. Polyn, and M. J. Kahana. Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological review*, 122(2):337, 2015.

[20] N. W. Morton and S. M. Polyn. A predictive framework for evaluating models of semantic organization in free recall. *Journal of Memory and Language*, 86:119–140, 2016.

[21] D. L. Nelson, C. L. McEvoy, and T. A. Schreiber. The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36(3):402–407, 2004.

[22] S. M. Polyn, K. A. Norman, and M. J. Kahana. A context maintenance and retrieval model of organizational processes in free recall. *Psychological review*, 116(1):129, 2009.

[23] L. Scrucca. GA: A package for genetic algorithms in R. *Journal of Statistical Software*, 53(4):1–37, 2013.

[24] P. Seitlinger, T. Ley, and D. Albert. Verbatim and semantic imitation in indexing resources on the web: A fuzzy-trace account of social tagging. *Applied Cognitive Psychology*, 29(1):32–48, 2015.

[25] S. Sen, S. K. Lam, A. M. Rashid, D. Cosley, D. Frankowski, J. Osterhouse, F. M. Harper, and J. Riedl. Tagging, communities, vocabulary, evolution. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 181–190. ACM, 2006.

[26] L. Steels. Collaborative tagging as distributed cognition. *Pragmatics & Cognition*, 14(2):287–292, 2006.

[27] E. Tulving. Episodic memory: From mind to brain. *Annual review of psychology*, 53(1):1–25, 2002.

[28] C. Wagner, P. Singer, M. Strohmaier, and B. A. Huberman. Semantic stability in social tagging streams. In *Proceedings of the 23rd international conference on World Wide Web*, pages 735–746. International World Wide Web Conferences Steering Committee, 2014.

[29] A. Zubiaga. Enhancing navigation on wikipedia with social tags. In *Wikimania 2009*. Wikimedia Foundation, 2009.