Balancing the Fluency-Consistency Tradeoff in Collaborative Information Search with a Recommender Approach

Paul Seitlinger, Tobias Ley, Dominik Kowald, Dieter Theiler, Ilire Hasani-Mavriqi, Sebastian Dennerlein, Elisabeth Lex & Dietrich Albert

Accepted author version posted online: 21 Sep 2017.
Published online: 21 Sep 2017.

Submit your article to this journal ☑

Article views: 15

View related articles ☑

View Crossmark data ☑

Taylor & Francis
Taylor & Francis Group

Balancing the Fluency-Consistency Tradeoff in Collaborative Information Search with a Recommender Approach

Paul Seitlinger[a], Tobias Ley[a], Dominik Kowald[b], Dieter Theiler[b], Ilire Hasani-Mavriqi[b,c], Sebastian Dennerlein[b], Elisabeth Lex[b], and Dietrich Albert[b,c,d]

[a]School of Educational Sciences, Tallinn University, Tallinn, Estonia; [b]Institute of Interactive Systems and Data Science, Graz University of Technology, Graz, Austria; [c]Know-Center GmbH, Research Center for Data-Driven Business & Big Data Analytics, Graz, Austria; [d]Institute of Psychology, University of Graz, Graz, Austria

ABSTRACT

Creative group work can be supported by collaborative search and annotation of Web resources. In this setting, it is important to help individuals both stay fluent in generating ideas of what to search next (i.e., maintain ideational fluency) and stay consistent in annotating resources (i.e., maintain organization). Based on a model of human memory, we hypothesize that sharing search results with other users, such as through bookmarks and social tags, prompts search processes in memory, which increase ideational fluency, but decrease the consistency of annotations, e.g., the reuse of tags for topically similar resources. To balance this tradeoff, we suggest the tag recommender SoMe, which is designed to simulate search of memory from user-specific tag-topic associations. An experimental field study ($N = 18$) in a workplace context finds evidence of the expected tradeoff and an advantage of SoMe over a conventional recommender in the collaborative setting. We conclude that sharing search results supports group creativity by increasing the ideational fluency, and that SoMe helps balancing the evidenced fluency-consistency tradeoff.

1. Introduction

Imagine a team of people in product development or organizational design with diverse backgrounds that have been given the task to deliver new kinds of products or solutions for designing new workspaces. They may try to research recent trends in workspace design to come up with innovative ideas. In today's work, such teamwork has become commonplace to deal with complex problems or for finding innovative solutions. In such a setting, it is necessary that the persons involved learn from one another, draw on their creativity, overcome groupthink (Janis, 1972; Page, 2007) and to come up with innovative product ideas (Paulus & Brown, 2007).

If the group uses information technology to do research and communicate, then this constitutes a networked search of solutions (e.g., Lazer & Bernstein, 2012), where each member receives help from human and non-human sources and contributes to a collective attempt to connect all those sources to a creative solution that is novel and useful. Digital curation, i.e., collaborating on the search and organization of problem-related sources (e.g., articles, videos, etc.) in social Web environments (e.g., Kerne, Smith, Koh, Choi, & Graeber, 2008; Kerne et al., 2014; Linder, Snodgrass, & Kerne, 2014), can be key in networked search. By raising one's awareness of others' contributions (e.g., collected sources) and reflections upon them (e.g., annotations), digital curation helps to mutually stimulate ideas "to increase one's potential for realizing

creativity" (Linder et al., 2014, p. 2411). In the sense of Sarmiento and Stahl (2008), digital curation supports the social dimension of creativity because it facilitates the building and maintenance of a shared problem space (e.g., of emerging relations between Web resources mediated by annotations) and bridging across different individuals' complementary ideas.

Lazer and Bernstein (2012) and Lazer and Friedman (2007) reported a number of theoretical and simulation-based studies in the tradition of networked search suggesting that effective search needs to balance divergent and convergent search processes. Divergent processes give each individual agent enough room for experimentation (e.g., autonomous exploration of information). From time to time, this needs to be balanced by convergent processes that allow exploiting and aligning each other's approaches toward a solution. The latter is usually accomplished by providing appropriate communication structures through which agents align their understanding.

Different research perspectives on social tagging (e.g., Fu, Kannampallil, Kang, & He, 2010; Lorince & Todd, 2016; Nelson et al., 2009; Pirolli & Kairam, 2012; Schweiger, Oeberst, & Cress, 2014) suggest that the use of tag-based annotations in digital curation could support such balancing. Tags are freely chosen keywords with which users describe resources on the Web and which may be visible to others. On the one hand, as social tags reveal other members' thoughts,

they trigger cognitive conflicts and inspire new ideas during individual experimentation (e.g., Schweiger et al., 2014). On the other hand, given sufficient consistency in applying certain tags for reoccurring topics, they support tag-based sharing of collected resources and facilitate an exploitation of own and others' search results (e.g. Fu et al., 2010; Lorince & Todd, 2016; Nelson et al., 2009; Pirolli & Kairam, 2012).

According to networked search (e.g., Lazer & Bernstein, 2012), a balanced view on the divergence–convergence continuum should lead to effective designs of digital curation environments. However, in social media and particularly social tagging studies, questions around the convergent pole have dominated the discourse and its agenda (e.g., Golder & Huberman, 2006). Probably because the lack of central control (e.g., standardized vocabularies) can lead to the "vocabulary problem" (divergent wording when tagging the very same object; Furnas, Landauer, Gomez, & Dumais, 1987), an endeavor of investigating and supporting consistency has come to the fore, for example by studying "semantic stabilization" (Wagner, Singer, Strohmaier, & Huberman, 2014). This focus on the convergent pole is also reflected by a large body of literature on the development of automatic tag recommendation mechanisms (TRM) (Dellschaft & Staab, 2012; Font, Serrà, & Serra, 2015; Jäschke, Marinho, Hotho, Schmidt-Thieme, & Stumme, 2007). TRM are services that encourage a convergent tag use and hence, alleviate the vocabulary problem by suggesting tags already applied in the past by other users. An example for a very simple strategy is represented by "most popular" recommenders which assume that what has been applied by many in the past is a good predictor for future assignments. Despite their simplicity, Most Popular Tag (MPT) recommenders work surprisingly well in predicting tag reuse in offline studies (Jäschke et al., 2007; Kowald et al., 2014).

In this article, we address the question of how to balance both processes in tag-based digital curation, i.e., how to increase exploration (divergent thinking of new search topics) against the backdrop of a sufficient level of tagging consistency in support of exploitation (making use of other persons' tags and associated search results). We investigate this question in a scenario where creative solutions are particularly important, namely in a work-integrated information search, and where a strong focus on convergent processes may especially be detrimental because persons are likely to share a common background and information goal, so that divergent processes need to be stressed to allow for creative solutions.

Rather than focusing on social imitation as many of the previous works have done in the area of tag-based curation, we approach this problem by drawing on the framework of "reflective search" (Seitlinger & Ley, 2016). The reflective search framework regards human web interaction as an iterative search of human memory shaped by past and present learning episodes. In our previous article, we have especially focused on convergent processes by looking at stabilizing tag vocabulary. In the present article, we draw our attention to the divergent pole of the exploration–exploitation continuum by considering effects of networked search on ideational fluency, a concept from the creative cognition literature (e.g., Benedek & Neubauer, 2013). In the present context, it

describes how easily and continuously diverse ideas can be accessed from memory during information-based ideation (Kerne et al., 2008, 2014), i.e., when thinking about search topics to be explored in future queries. Referring to previous work on cognitive effects of social tags on mental structures (e.g., categories and associations; e.g., Fu & Dong, 2012; Seitlinger & Ley, 2012; Seitlinger, Ley, & Albert, 2015), we anticipate a tradeoff between fluency and consistency: when users of a digital curation environment perceive others' tags, these tags leave episodic memory traces (Seitlinger & Ley, 2012; Seitlinger et al., 2015), strengthening previously weak associations to a search topic, in case these traces represent new ideas. Considering research in creative cognition (e.g., Smith, Ward, & Finke, 1995; Ward, 2007), this tag-based cognitive effect should reduce the dominance of pre-existing stereotyped associations and give rise to a broader (mental) fan of equally available ideas around a topic. This is also called a flatter hierarchy of associative strengths (Benedek & Neubauer, 2013; Mednick, 1962). While such a mental organization allows for a steadier stream of ideas, i.e., more fluency, the increased availability of several responses to a certain topic should simultaneously decrease tagging consistency.

The first research question of this work is *whether the anticipated fluency-consistency tradeoff can be evidenced when people perform tag-based search and curation collaboratively (seeing each other's tags and search results) in contrast to when they perform this search individually*. In particular, we seek to demonstrate the flattening impact of social cues (e.g., tagged bookmarks) on people's associative hierarchies by the manifestation of the tradeoff (decreased consistency and increased fluency) in a realistic information search scenario at the workplace. We let persons (research staff) bookmark and tag Web resources on a given topic ('redesigning workspaces to move people') both under an individual and collaborative search condition. Under both conditions, we then examine each person's tagging consistency (extent of choosing similar tags for topically similar bookmarks) and ideational fluency (stream of responses in a free association task on a set of subtopics, such as 'interior design', 'inspiration sources', etc.) and expect the experimental variation (individual vs. collaborative) to lead to increases in fluency being accompanied by decreases in consistency, and vice versa.

In the light of the previously mentioned studies of Lazer and Bernstein (2012) and Lazer and Friedman (2007) on the tradeoff between divergent and convergent search processes in effective problem solving in groups, this research question is not novel. Moreover, in this context, studies on creative group cognition (e.g., Kohn, Paulus, & Choi, 2011; Nijstad & Stroebe, 2006) should also be mentioned, which have already found evidence that mutual stimulation in collaborative brainstorming settings can have positive effects on individual fluency scores. Therefore, investigating our first research question is primarily of an incremental nature and asks whether existing evidence gathered under controlled experimental conditions generalizes to more natural conditions of an information search at the workplace. Especially, from an applied perspective, such validation appears important to us because it would imply that designing social systems for

information search can be directed either toward convergence (i.e., tagging consistency), or toward divergence (i.e., ideational fluency). The former would aim for aligning the vocabulary (e.g., Font et al., 2015), while the latter for stimulating creation (e.g., Candy & Hewett, 2008), and it would seem difficult to balance the two complementary processes. Considering recent discussions around filter bubbles in Web environments (e.g., Pariser, 2011; Sunstein, 2001), we consider such research to contribute to a more balanced view on interaction in the social web.

As a second aim of this article, we therefore address this design challenge by introducing a tag recommendation mechanism (TRM) that compensates for the hypothesized downside of decreased consistency by pushing the reuse of tags for certain topics, even if associative hierarchies are flat. To design such a TRM, we draw on formalisms of memory psychology that help translate the reflective search framework into a computational model to be readily applied as a TRM. As summarized in Figure 1, the first step of translation, i.e., a formalization of the model, makes use of a stochastic account of human memory search (e.g., Unsworth & Engle, 2007) to specify cognitive structures (e.g., an associative hierarchy) and processes (e.g., encoding and retrieval), which are involved in reflecting on objects and underlie the tradeoff between fluency and consistency. The second translation step then instantiates the search-of-memory formalization to create a TRM, which emulates a person's associations by tracking and consolidating tag choices for particular resource topics. Based on this memory-like representation of tagging behavior, the mechanism is able to mimic a resource-triggered search of memory (SoMe) for topically relevant tags. This should confer an advantage over conventional most popular tags (MPT) approaches, if associative hierarchies are flat. *Therefore, the second research question is whether SoMe achieves high tag acceptance rates under a collaborative search where* associative hierarchies are assumed to be flat, and, more specifically, whether the SoMe advantage (over MPT) is larger under the collaborative than individual search condition.

Summarizing, while our first research question asks for empirical evidence of the fluency-consistency tradeoff, the second points toward an effective strategy to alleviate inconsistent tagging behavior in case ideational fluency is high. We will now turn to the first research question and present an empirical study conducted to that evidence. In the article's second part, we will then tackle the second question and introduce an effective recommender approach that is applied in the same study.

2. Evidence of the fluency-consistency tradeoff in collaborative information search

2.1. Background and hypotheses

As the question of the fluency-consistency tradeoff is built upon specific assumptions on the mental organization of a person's associations (the associative hierarchy), we first provide a more formal and process-oriented interpretation of an associative hierarchy in form of a stochastic model of memory search. Based on this model, we specify the assumed effect of social cues (during a collaborative search) on an associative hierarchy and derive the hypotheses on the fluency-consistency tradeoff.

A process-oriented interpretation of an associative hierarchy

Our starting point of formalizing the associative hierarchy is the reflective search framework (Seitlinger & Ley, 2016). We assume that reflections upon objects (e.g., when tagging topics of an article, or generating ideas to approach a problem, etc.) are accompanied by a search of secondary memory. Referring
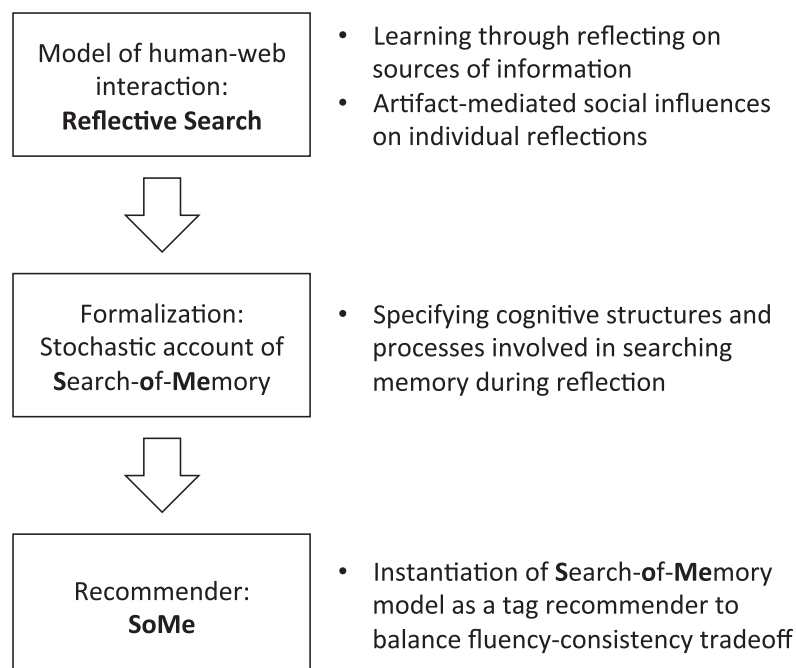

Figure 1. Design process translating reflective search model into a service that balances the fluency-consistency tradeoff.

to long-standing research on human memory search (for a review see Davelaar & Raaijmakers, 2012), this process is triggered by an environmental cue, such as a problem to be solved or an article to be tagged. This cue is assumed to activate a mental search set S (see Figure 2), a reservoir of associations, from which a number of N targets (e.g., problem-relevant ideas, or topic-related aspects; schematized by black filled dots) can be sampled (i.e., brought to mind) with a particular search rate λ.

The temporal dynamics of memory search are driven by an inverse relationship between N and λ (e.g., Albert, 1968; Bousfield, Sedgewick, & Cohen, 1954; Kaplan, Carvellas, & Metlay, 1969): If many ideas about a given problem or aspects of an article's topic can come to mind (large N; right search set in Figure 2), the search rate is reduced (small λ) due to more competition between the available ideas – with "each [idea] competing against all of its peers" (Rohrer, 2002). Put differently, the mental search set is in a state of defocused activation (e.g., Dorfman, Martindale, Gassimova, & Vartanian, 2008; Martindale, 1995) that is distributed among several elements with comparatively low relative strengths. Such an activation state constitutes a flat associative hierarchy allowing for a slow but steady stream of ideas (Mednick, 1962). To the contrary, if only few ideas have strong associations to a given topic, the search set exhibits a state of focused activation shared among only few associations with high relative strengths (left search set in Figure 2). The resulting steep hierarchy becomes manifest in a fast retrieval (large λ) of only few ideas (small N).

As mentioned above, one way to reveal a person's associative hierarchy, i.e., to estimate N and λ, is to measure ideational fluency with respect to the topics of the collaborative

information search. To this end, we drew on a free association task, and a corresponding stochastic model to analyze the responses and derive the two parameters. In the free association task, a participant is exposed to a cue (e.g., the topic 'interior design') and asked to name as many associations (ideas) that come to mind. The triggered stream of responses is recorded in form of the cumulative number of unique ideas (ideational fluency) and then analyzed in terms of the temporal dynamics, i.e., inter-response times (in seconds; see schematic diagrams on the bottom of Figure 2). In the present study, we have used the topics of the information search (e.g., 'interior design', 'augmented reality', etc.) as cues for a Web-based free association task at the beginning of the study and after the individual as well as collaborative search condition.

In order to derive estimates of N and λ from the recorded ideational fluency, search through memory can be approximated as a random search process following a repeated sampling-with-replacement scheme (e.g., Bousfield & Sedgewick, 1944; Wixted & Rohrer, 1994; Unsworth & Engle, 2007). At the beginning, i.e., on the first sample, S includes N targets (relevant ideas) and S – N non-targets. Since sampled targets are replaced, the rate of producing new associations, λ_i, decreases linearly with the number of already sampled targets i according to $\lambda_i = (N - i)\lambda$ (e.g., Albert, 1968). The consequence is an exponential decay function (Bousfield & Sedgewick, 1944) given by

$$F(t) = N\left(1 - e^{-\lambda t}\right) \tag{1}$$

where F(t) represents the cumulative number of unique responses by time t. A large number of studies demonstrates that the cumulative exponential of Equation (1) can be fitted to the response protocol gathered by a free association task to
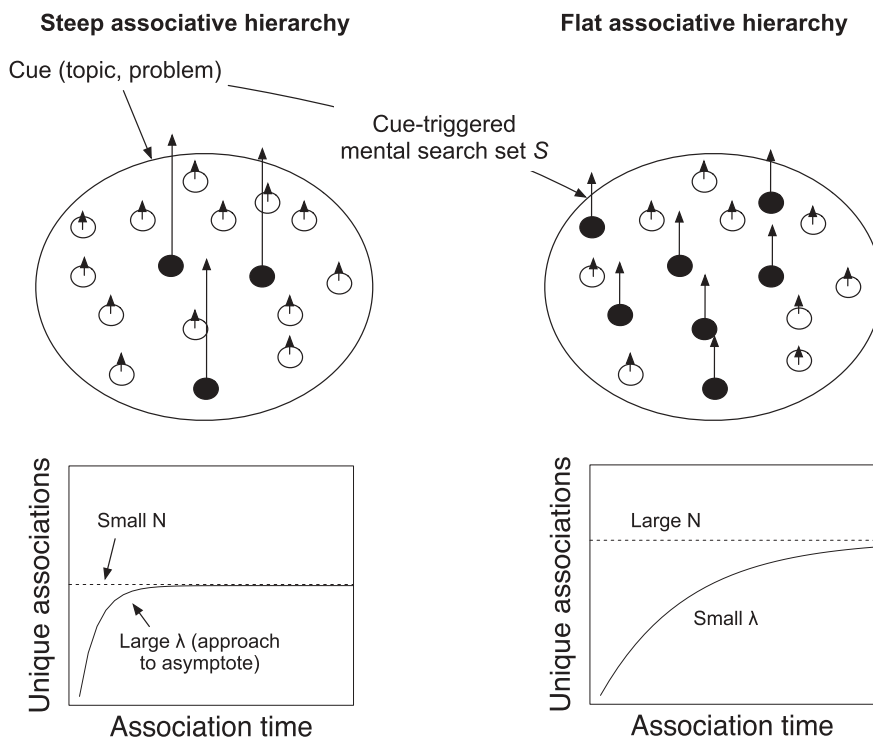

Figure 2. Associative hierarchy as a mutual dependence of N (asymptotic number of topic-related associations) and λ (rate of approach to N).

estimate N and λ (e.g., Wixted & Rohrer, 1994). In the present study, we have adopted this approach to characterize the associative hierarchies of the participants with respect to the topics explored during their information search.

Next, we make use of this stochastic search scheme to specify the effect of social cues (e.g., tagged bookmarks) on a person's associative hierarchy and to derive our hypotheses on the fluency-consistency tradeoff.

The impact of social cues on an associative hierarchy

When people collect and tag bookmarks of resources on the Web, we assume their thoughts and ideas to develop within an evolving practice that is co-created in a collective and artifact-mediated activity (e.g., Hutchins & Johnson, 2009). As a methodological consequence, we do not aim to decontextualize our unit of analysis, i.e., associative hierarchy, under laboratory conditions, but try to shed light on its relations to the natural and cultural environment and practice, in which it evolves and – at the same time – to which it contributes (e.g., Roepstorff, Niewöhner, & Beck, 2010).

To attain a holistic picture of the evolving practice, we also consider the active role of non-human actors (Latour, 2005), such as tag clouds or recommender mechanisms, which mediate the co-creation of environmental structures by distributing joint artifacts (e.g., social tags) and thus, affecting mental structures (e.g., Schweiger et al., 2014). For instance, we assume that a person's mental search set S (the set of associations activated by a given topic) is affected by joint artifacts (tags that have been introduced by other people and become visible through a tag cloud or recommendation mechanism). Evidence for the assumed long-term influence of those tags on mental associations comes from studies demonstrating perceived tags to leave robust memory traces and to affect future tag choices (e.g., Seitlinger et al., 2015). Given that some of these experienced tags convey new and interesting ideas the

person hasn't thought of before, the number N of topic-related associations (targets) should therefore increase over time within the shared bookmarking system. The assumption that tags help a person experience new ideas cannot be observed directly. However, we think the assumption is warranted, if active engagement with others' tags can be observed (e.g. through analyzing the log-file), and at the same time estimates of N are in fact larger after a collaborative than individual search. In the results section, we will be offering some more insights on whether tag-mediated mutual stimulations has likely happened in our case.

Figure 3 illustrates the expectation of tag-based influences on a person's mental search set S and contrasts a collaborative with an individual search condition. The tag clouds in the shared environment as well as the web resources others have contributed should expose a person to different perspectives on a given topic, resulting in more topic-relevant associations (dark shaped dots) within S. The consequence should be a flattening of the associative hierarchy: activation among elements in S should become defocused (larger N) and each association's relative strength should decrease and become less available (smaller λ). On the other hand, the activities in the individual bookmarking condition give rise to tag clouds resulting exclusively from an individual's contributions. In this case, the function of the tag clouds is less to propagate new than to reinforce existing associations, resulting in a smaller increase of new topic-relevant associations. This 'individualistic' interplay of environmental and mental structures mediated and reinforced by tag clouds should thus give rise to a comparatively steep associative hierarchy: a focused activation within S (small N) accompanied by a fast access to the correspondingly few associations (large λ).

As already stated, we assume the shape of an associative hierarchy (steep vs. flat) to have opposing effects on ideational fluency on the one hand, and tagging consistency on the other
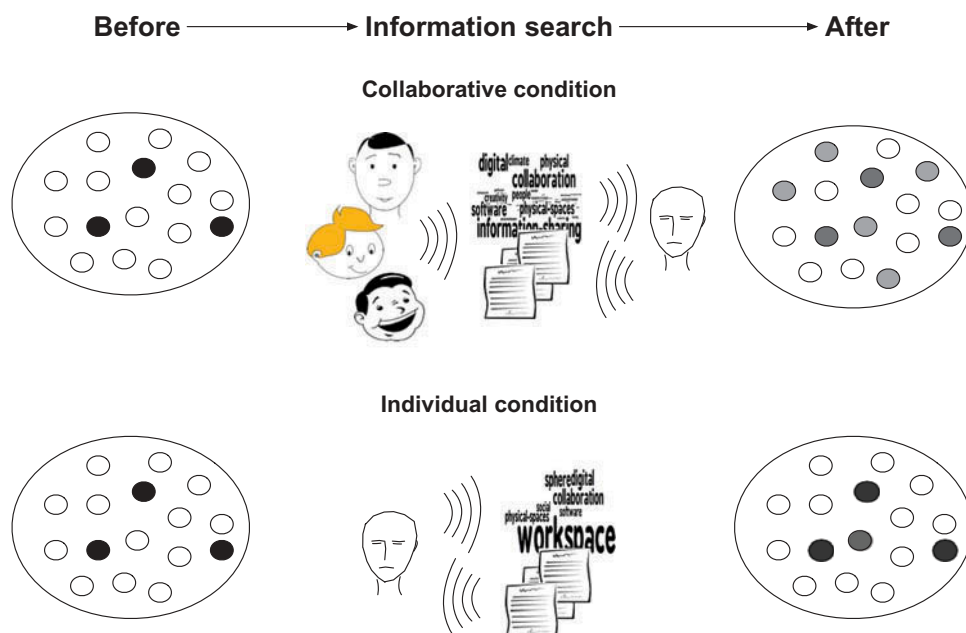

Figure 3. Artifact-mediated mutual stimulation causes associative hierarchy to be flatter after a collaborative than individual information search.

hand. This fluency-consistency tradeoff is described next as well as the two hypotheses following from it.

Testing the fluency-consistency tradeoff: Hypotheses H1.1 and H1.2

Ideational fluency was measured by means of the stream of responses in a free association task characterized by N (the asymptotic number of associations) and λ (the rate of approach to the asymptote). Based on a number of studies on retrieving from semantic memory (for a review see Wixted & Rohrer, 1994), we assumed the following relationship: The flatter the hierarchy, the larger the estimate of N and the smaller the estimate of λ (see Figure 2). In the current study, each participant performed the task three times: at the beginning of the study (baseline measurement), after an individual and after a collaborative information search (counterbalanced repeated measurement). The stimuli were eight sub-topics of the search task ('augmented reality', 'health', 'interior design', 'gamification', 'inspiration sources', 'collaboration technologies', 'personalization services', 'socializing'; for details see Section Search task and bookmarking interface) and held constant across the three points of measurement. Based on the assumption that the collaborative condition results in flatter associative hierarchies than the individual condition, *the first hypothesis was that relative to the baseline measurement, participants exhibit a stronger increase in estimates of ideational fluency (larger N and smaller λ in the free association task) after the collaborative than after the individual search (H1.1).*

Tagging consistency, the second indicator applied to characterize an associative hierarchy, was defined as the extent to which similar tags were assigned to bookmarks of semantically similar resources, i.e., resources dealing with similar topics. With respect to its relationship to the associative hierarchy, we assumed that if several associations compete for indexing a topic (flat hierarchy), the tagging behavior for resources of that topic should be more variable than if only few associations have high probabilities being retrieved (steep hierarchy). To quantify consistency, we implemented a tagging interface that prompted a person to describe each bookmark semantically (by selecting from a list of the eight search topics) and by a set of freely chosen tags. Thus, for two bookmarks x and y, we could calculate a topical similarity score S_T (normalized topic overlap of x and y) and a verbatim similarity score S_W (normalized tag overlap of x and y). Finally, we defined tagging consistency $r(S_T, S_W)$ as the correlation between S_T and S_W across all pairs of collected bookmarks separately for the individual and collaborative search condition. See Section Measures and statistical analysis for more details on the two similarity scores and calculating $r(S_T, S_W)$. Based on the assumption that the collaborative condition results in flatter associative hierarchies than the individual condition, *the second hypothesis was that estimates of $r(S_T, S_W)$ are larger under the individual than collaborative information search (H1.2).*

Summarizing, the two hypotheses H1.1 and H1.2 specify the expected fluency-consistency tradeoff that is mediated by the shape of the associative hierarchy. While a flat hierarchy favors ideational fluency, a steep one brings forward its conceptual counterpart, i.e., tagging consistency.

2.2. Method

We investigated the hypotheses H1.1 and H1.2 within the scenario of a work-integrated information search on the topic 'workspace redesign to move people', which was supported by a bookmarking system for collecting (bookmarking) and tagging topic-relevant Web resources. Our expectation was that a collaborative search condition (shared bookmarking system) on average results in a flatter associative hierarchy than an individual condition (unshared system) and that this difference manifests itself in a fluency-consistency tradeoff: a higher ideational fluency (H1.1) and a lower tagging consistency (H1.2).

Participants

The information search was performed by $N = 18$ researchers ($n = 6$ female) from different groups across different institutions, with an average age of 31.5 years ($SD = 5.5$, ranging from 23 to 46 years). The following research groups participated in the study: one cognitive science group at an Austrian technical university ($n = 6$), two groups from an Austrian research institute dealing with social computing ($n = 7$) and ubiquitous computing ($n = 3$), and one group on educational technology from an Estonian university ($n = 2$). All research groups were interdisciplinary having members on computer science, humanities, and psychology.

Particular measures were taken to ensure that persons of a shared bookmarking system could only influence each other via shared artifacts (i.e., bookmarks and tags) independent of their real geographical distance or research group membership: First, the assignment of the participants to the experimental conditions was random and did not take into account the research group membership. Second, every participant was visible in a bookmarking system only in form of a pseudonym, which was drawn randomly from a pool of popular English names (e.g., George) and did not allow inferences on her or his real name or identity. Third, participants were instructed not to discuss ongoing activities within the bookmarking system with other participating colleagues, where on average, the probability of sharing a system with a colleague from the same research group was $p = .14$. As furthermore all resources and tags had to be in English, we were confident that having two participants contribute from Estonia and the remaining sample from Austria had no significant influence on the activities going on in the bookmarking system.

Design

The independent variable, denoted 'Search Condition', differentiated between a collaborative and an individual information search. The latter took place in a separate bookmarking system only displaying each employee's own Web resources (in form of a list) as well as her/his own tags (in form of a tag cloud). Under the collaborative condition, the employees shared a social bookmarking system making available the resources and tags of all the system's members. To increase statistical power, we realized a randomized counterbalanced repeated measurement design: Every employee collected Web resources under both conditions for two weeks each, where one half of the participants switched from the individual to

the collaborative and the other half from the collaborative to the individual condition. For statistical analyses, we then merged the data of the two individual and the two collaborative study halves.

The dependent variables were (a) ideational fluency (measured by the Web-based free association task each employee performed at the beginning, after the individual, and after the collaborative search condition), and (b) tagging consistency (determined by a log file-based analysis on the correlation between the topical similarity S_T and the tag similarity S_W of an employee's bookmarks). The design included one additional independent variable (the type of Tag Recommendation Mechanism displayed), as well as one additional dependent variable (acceptance of these recommendations). We will cover this part of the design in the article's second part dealing with the tag recommendation mechanism SoMe.

Search task and bookmarking interface

The information search had the character of a simulated workplace learning scenario as the search topic 'workspace redesign to move people' was not part of an ongoing research project but defined specifically for the purpose of the study. However, to make it as realistic and motivating as possible, the topic was co-defined together with the work group leaders as a topic they expected to stimulate valuable workflow reflections and improvements. Insights gained during search were therefore discussed subsequent to the study in the context of work group meetings. While the specific search environment that was used was new to the participants, the way the search task was set up (e.g., collecting resources and sharing them in an online system) corresponded to how typically explorations of new topics were done in these institutions.

Instruction. Instructions and passwords (to enter the bookmarking system) were sent via e-mail. The instruction described the search topic, which was '*workspace (re)design to move people – improving knowledge exchange and creation in your work group*'. For the coming four weeks, the employees were asked to imagine the task of writing a state of the art for a project proposal that '*sheds light on the topic from different perspectives*'. To this end, each employee had to bookmark at least three to four Web resources (e.g., articles, videos) per week and to annotate each bookmark '*by means of predefined topics (e.g., 'inspiration sources & techniques') and freely chosen tags (e.g., 'physical_proximity', 'random_encounters', etc.)*'. Under the collaborative condition, the employee was also instructed to attend to each other's contributions (tags and shared bookmarks) to get to know different perspectives on the task.

Bookmarking interface. A Web resource was bookmarked by means of an interface displayed in Figure 4, which prompted an employee to annotate a resource by choosing one or several topics from a predefined eight-item list (number 2), and assigning tags by choosing from a list of recommendations (number 3) and/or typing in personal tags (number 4). The subset of chosen topics, denoted T, was logged for a later analysis of the employee's tagging consistency as well as to trigger the presentation of the set of
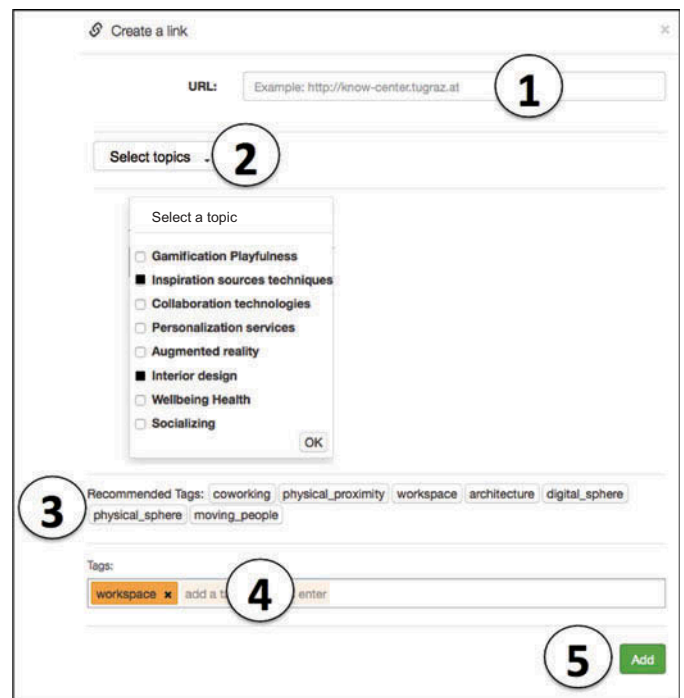

Figure 4. Bookmarking interface to collect a Web resource (number 1), classify it by choosing from a list of pre-defined topics (number 2), receive a set of recommended tags, denoted W_{REC} (number 3), and make a tag assignment (number 4).

recommended tags. It was therefore important to provide a list of topics, which, from the viewpoint of the employees, covered important thematic aspects of the search task and was readily comprehensible for them based on their prior knowledge. To gather such a list, a Web-based idea generation task had been administered one week before study start, asking each employee to "list as many design ideas as possible for a workspace, which could improve the exchange and creation of knowledge in your work group". All listed ideas were then subjected to a qualitative content analysis identifying the most important workspace dimensions mentioned by the employees and reducing these dimensions to the following eight topics: 'Interior design', 'Inspiration sources & techniques', 'Collaboration technologies', 'Gamification & Playfulness', 'Personalization services', 'Augmented reality', 'Wellbeing & health', and 'Socializing'.

Though this list was certainly incomplete, it was exhaustive with respect to the responses the present sample of participants had produced in the idea generation task. To further take account of the opportunity that new topics can be discovered during search, we instructed the participants to inform us whenever they would become aware of an important topic not included in the current list. As we, however, never received such feedback from the participants, the eight-topic list depicted in Figure 4 (number 2) was maintained during the whole study duration.

Measures and statistical analysis
Free association task to observe ideational fluency. To measure ideational fluency, every employee generated free associations for 60 s to each of the eight topics in a Web-based free association task (FAT) at the beginning

(baseline measurement) and after each of the two search conditions (individual and collaborative). Per employee and per point of measurement (baseline, individual, collaborative), we gathered an average FAT response protocol, which was the cumulative number of associations per second averaged across the eight stimuli (topics). In a final step, we merged the average protocols of all 18 employees to characterize the average ideational fluency under each of the three points of measurement in terms of N and λ (by fitting the cumulative exponential of Equation 1).

Tagging consistency. To measure each employee's tagging consistency, we first extracted from the log data all the bookmarks that s/he had collected, where each bookmark was characterized semantically by the set of selected topics (e.g., 'interior design', 'inspiration sources'), denoted T, and by the set of assigned tags (e.g., 'creativity', 'curation', 'mental_fixation') denoted W. Then, for each pair of the employee's bookmarks we calculated two scores, a topic similarity score S_T and a tag similarity score S_W, by applying the *Jaccard* index. Thus, for any two bookmarks x and y, S_T and S_W were given by

$$S_T(x,y) = |T_x \cap T_y| / |T_x \cup T_y| \qquad (2)$$

,

$$S_W(x,y) = |W_x \cap W_y| / |W_x \cup W_y| \qquad (3)$$

E.g., if T_x is {'inspiration sources & techniques', 'collaboration technologies'} and T_y is {'collaboration technologies', 'personalization services'}, the intersect and union of both sets include one and three elements, respectively, resulting in $S_T = 1/3$. Above, we defined tagging consistency as the correlation between S_T and S_W. Therefore, to quantify the predicted differences in consistency between the individual and collaborative condition, we performed a regression of S_W on the continuous predictor S_T and the categorical predictor 'Search Condition', and included an interaction term to test the predicted assumption of different slopes under the individual and collaborative condition.

2.3. Results

Our research design and methodology aimed at investigating the assumption that an individual's associative hierarchy becomes flatter during interactions with joint artifacts (i.e., social tags) that are propagated in a shared bookmarking system by virtue of tag clouds. Based on this assumption, we expected a fluency-consistency tradeoff that becomes manifest in a higher ideational fluency (H1.1) and a lower tagging consistency (H1.2) under the collaborative than individual information search.

Hypothesis H1.1

Our first hypothesis H1.1 was that employees exhibit a higher ideational fluency in a free association task (FAT; larger N and smaller λ) after a collaborative than individual information search. Figure 5 presents the average ideational fluency data – cumulative associations (workspace design ideas) generated
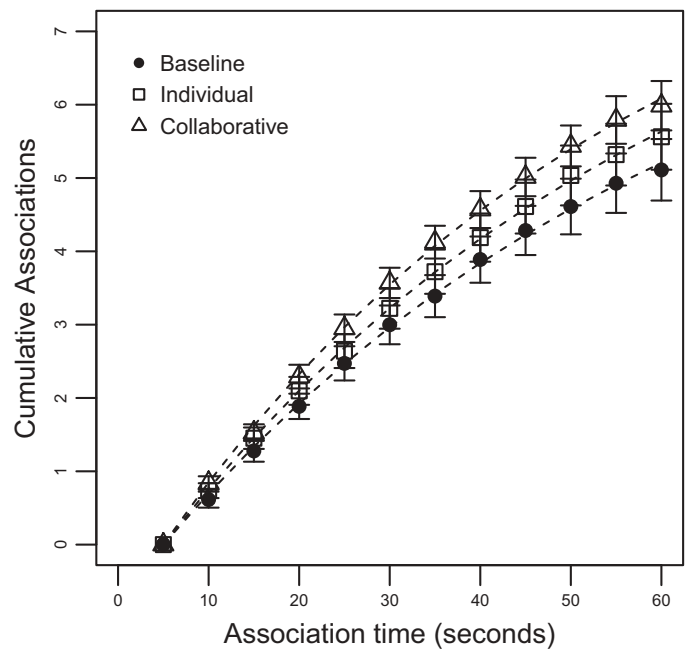

Figure 5. Cumulative free association latency distribution. Error bars represent 1 standard error of mean.

by 18 employees to eight different topics as a function of time (seconds) – before the study start (baseline condition; black-filled circles), after the individual search condition (squares), and after the collaborative search condition (triangles). A glance at the three latency distributions reveals a general learning effect of the search task: in comparison to the baseline distribution, employees appeared to produce more associations both after the individual and collaborative search. In addition and according to our expectation, this learning effect seemed to be larger under the latter condition.

In line with research on retrieval from semantic memory (e.g., Rohrer, 2002), the model-based analysis showed that the rate of producing new associations slowed continuously in time and could be well described by the cumulative exponential given by Equation 1. The best-fitting parameter estimates of the exponentials (dashed lines), together with the percent of variance explained, are presented in Table 1. With respect to parameter N, the estimates lend support to Hypothesis H1.1 that this asymptotic number of produced associations increased monotonously from the baseline, over the individual, up to the collaborative condition. Thus, the employees could increase their knowledge especially in the course of the collaborative search, i.e., they could substantially extend the pool of relevant ideas about the eight search topics.

However, in contrast to our expectation, this monotonous increase was not related inversely to the parameter λ, the rate of approach to asymptote, whose estimates were approximately equal for the baseline and individual and, in fact, largest for

Table 1. Best-fitting parameter estimates of association latency distributions.

Condition	N	λ	Variance explained
Baseline	8.610	.086	.66
Individual	9.156	.087	.67
Collaborative	9.388	.095	.81

the collaborative condition. That is, the collaborative search resulted in an increase of both the number and speed of produced associations. In the light of the stochastic search model, this pattern might be explained post-hoc by the assumption that the activation in the mental search set S was not only spread among a larger number of N targets (relevant ideas) but also drained out of the remaining S–N non-targets (irrelevant associations) e.g. through stronger lateral inhibition. In other words, the learning process during collaborative search allowed, on the one hand, getting to know more ideas and on the other hand, effectively inhibiting irrelevant associations or perhaps even replacing them from S. Note that such a pattern still implies a mental activation that is distributed more evenly among a larger number of targets (topic-relevant associations) and that the collaborative search brings about a flatter associative hierarchy for relevant ideas than the individual search.

Finally, a Friedman test on differences in the number of unique associations among the three conditions of baseline (Median = 4.81), individual search (Median = 5.00), and collaborative search (Median = 5.69) reached significance at the .05 level, $\chi^2(2, N = 18) = 5.68$, $p = .06$. Pairwise comparisons using Wilcoxon test and controlling for the Type I errors by using the LSD procedure further revealed that this effect could be attributed to the difference between the collaborative and baseline condition, $p < .05$; there were no differences neither between the collaborative and individual, $p = .46$, nor between the individual and baseline condition, $p = .45$. In other words, only the collaborative search gave rise to fluency scores that contrasted significantly with those scores that the participants had already been able to achieve before performing the information search.

Summarizing, hypothesis H1.1 assumed ideational fluency to be greater after the collaborative than individual search. As the employees exhibited both the largest number and highest speed of responses under the collaborative condition, we interpret the results as providing support to H1.1. In particular, they harmonize with the assumption of mutual stimulation through joint artifacts, i.e., social tags that act as sign vehicles propagating diverse ideas among the employees. This assumption was further corroborated by a descriptive analysis of participants' click behavior, in particular, their clicks on tags in the shared tag cloud to filter already collected bookmarks within the system. The analysis showed that the probability of a person clicking on a tag that had been introduced by a different person was relatively high, i.e., $p = .68$ ($SD = .27$), indicating curiosity-driven search behavior and the intentional use of tags to discover novel sources of information. In light of this additional pattern, the active role of tags in mediating the experience of novel ideas appears even more likely. From a learning perspective, this tag-based propagation of ideas is highly desirable, as it seems to broaden and flatten the hierarchy of topic-related associations and thus, supports a more creative encounter with a given topic (e.g., Benedek & Neubauer, 2013).

At the same time, however, a broader and, in addition to that, easily accessible hierarchy of topic-related associations around a given topic can be expected to result in a stronger variability in tag choices for related Web

resources. In the following, we therefore investigate this tradeoff, i.e., the hypothesized downside of a flatter associative hierarchy, namely a higher tagging inconsistency (hypothesis H1.2).

Hypothesis H1.2

Our second hypothesis was that a flatter associative hierarchy under the collaborative condition should become manifest in a more inconsistent tagging behavior, i.e., in a weaker tendency to assign similar tags to topically similar Web resources. In particular, H1.2 was that the relationship between S_W and S_T is stronger under an individual than collaborative search condition. To test H1.2, we first gathered all pairs of an employee's bookmarks, determined each pair's tag and topical similarity (S_W and S_T, respectively; see Equations 2 and 3) and finally, performed a regression of S_W on the continuous predictor S_T and the categorical predictor 'Search Condition'.

685 data points entered the regression, which consisted of the 361 bookmark pairs of the individual and of the 324 bookmark pairs of the collaborative condition. The model explained about 13% of variance in the tag similarity S_W of an employee's pair of bookmarks (adjusted $R^2 = 0.132$, $p < .001$) and yielded a highly significant effect for the predictor S_T ($t = 9.87$, $p < .001$), and – in line with H1.2 – a highly significant interaction between this continuous predictor and the categorical predictor 'Search Condition' ($t = -4.44$, $p < .001$). Table 2 shows the estimates of the model's intercept and slope and how these estimates change as a function of 'Search Condition'. The small amount of variance explained is not surprising as the probability of reusing tags (that underlies S_T) depends not only on semantic attributes of the resources, but also on mediating cognitive processes (e.g., Fu & Dong, 2012; Seitlinger et al., 2015), which are to some extend specified by the tag recommender's algorithm presented in the article's second part. The present regression model, however, did not capture such cognitive processes because its primary purpose was not to explain a large amount of variance in the individuals' tagging behavior but to determine whether the amount of variance explained by the predictor S_W differs between the individual and collaborative condition.

Under the individual information search, the standardized coefficient β_1 (i.e., the slope of the predictor S_T) takes on the value of about 0.40, which is indicative of a moderate effect size. Thus, the higher S_T, i.e., the higher the topical similarity of two Web resources collected by an employee, the higher is the similarity of the tags S_W applied to that pair of Web resources. However, as Table 2 also shows, the predictor's slope significantly declines under the collaborative condition and, in fact, takes on the value of only 0.14, which is indicative of a small effect size. Thus, the small amount of variance explained (ca. 13%) in the whole dataset can probably be attributed to the fact that S_T is a substantial predictor of S_W only under the condition

Table 2. Summary of the regression of S_W (tag similarity of a bookmark pair) on S_T (the pair's topical similarity) and 'Search Condition'.

Search condition	Intercept	Slope
Individual	$\beta_0 = 0.05$	$\beta_1 = 0.39$
Collaborative	$\beta_0 + \alpha = 0.12$	$\beta_1 + \beta_2 = 0.14$

Note. $\alpha = 0.07$; $\beta_2 = -0.25$.

of an individual information search. These results are well in accordance with H1.2, which assumes that under the collaborative condition, where tag clouds propagate divergent tag-topic associations, the tendency to reapply the same tags for reoccurring topics (semantically similar Web resources) is smaller than under the individual condition.

Controlling for priming effects.[1] Especially with respect to hypothesis H1.2 (on differences in participants' tagging consistency), a question may be to what extent our result patterns were an artifact of having participants select from the eight pre-defined topics, as this potentially had priming effects on their subsequent choice of tags. However, as our main interest was in investigating the impact of an experimental variable (differing search conditions), we did not expect our results to be sensitive to the topic structure because topic-based priming should not vary between the conditions. Nevertheless, in order to explore its potential impact, we performed a statistical control analysis, which is described next.

We performed a second regression analysis by entering the categorical predictor 'topic use'. Based on a median split, 'topic use' distinguished between participants who on average selected many vs. few topics to describe a given bookmark. The rational behind it was that if topic-based priming effects were actually negligible for our results (on condition-dependent differences in consistency), hypothesis H1.2 (stronger $S_T - S_W$ relationship under the individual than collaborative condition) should hold in both pre-experimental groups that can be assumed to be affected by different priming effects due to different ways of interacting with the topic structure.

The dummy variable split the sample along the median of 1.8 into a "few topics" and a "many topics" group that on average had selected 1.44 and 2.46 topics per resource. This extended regression model indeed yielded a second-order interaction ($t = 2.51$, $p < .05$), which, however, was of an ordinal nature and only qualified our previous data interpretation: under both groups, the relationship between S_T and S_W was stronger under the individual than collaborative search, where this difference was greater in the "few topics" (standardized coefficients: $\beta_{individual} = .42$, $\beta_{collaborative} = .05$) than "many topics" group ($\beta_{individual} = .31$, $\beta_{collaborative} = .23$). This effect can be explained by drawing on the search of memory model (Figure 2): selecting more topics provides more context cues, which constrain the mental search set more strongly and in further consequence, reduce the variance in tagging behavior across time and different environmental conditions, such as those of an individual vs. collaborative search. Summarizing, the regression model extended by the variable of 'topic use' further strengthened hypothesis H1.2 because the experimentally induced difference in consistency could be observed in both pre-experimental groups. We therefore conclude that topic-based priming processes did not contribute substantially to the observed differences in search condition. Hence, our results should generalize to other settings that e.g. make use of a more or less differentiated topic structure.

2.4. Discussion

We gathered evidence for the expected fluency-consistency tradeoff by observing opposing effects of 'Search Condition' on temporal dynamics of free associations and tagging behavior: Displaying social cues (under the collaborative condition) increased ideational fluency (measured by a FAT), but reduced the consistency of tag choices for reoccurring topics (measured by a correlational analysis of log data). Networked search (e.g., Lazer & Bernstein, 2012), however, benefits from balancing individual exploration – a process that benefits from ideational fluency (e.g., Kerne et al., 2014) – and exploitation of each other's search results, which, in the context of tag-based bookmark sharing, is facilitated by topically consistent tag choices (Dellschaft & Staab, 2012). Therefore, finding evidence of the fluency-consistency tradeoff begs the question of how we are to deal with the downside of a flat associative hierarchy, i.e., growing inconsistency.

One potential answer will be explored in the next section where we have contrasted the use of two tag recommendation mechanisms (TRM). One of those mechanisms is built on a "Most Popular Tag (MPT)" recommendation strategy, a variant of which can be found in many contemporary web environments. The other is derived from a mechanism that models associative hierarchies over different search topics and produces tag choices as a search of memory. As the latter is especially tuned to a situation of flat associative hierarchies and increased ideational fluency, it should be especially effective in the collaborative information search condition.

3. An effective recommendation mechanism for collaborative information search

Tag recommendation mechanisms (TRM) can be seen as non-human actors (Webster, Gibbins, Halford, & Hracs, 2016) to which developers of social information systems delegate the intention of making information search more effective. TRM can be directed either toward convergence (i.e., tagging consistency) for the sake of aligning the vocabulary (e.g., Font et al., 2015), or divergence (i.e., ideational fluency) for stimulating creative ideation (e.g., Kerne et al., 2014). For system designers, this means it is difficult to achieve both things at the same time.

When designing TRM, we draw on our framework of reflective search. By modeling a fundamental search process, namely that of search of (human) memory when reflecting on a resource, it allows to deal with cognitive dynamics involved in the tradeoff between exploration and exploitation of networked search. We introduce SoMe, a TRM that simulates search of memory dynamics in response to a resource and hence, to identify tags that are co-created in collaborative search and are likely to resonate with a user's current reflection. That way, the probability of reusing tags for re-occurring topics (i.e., tagging consistency) can be increased and the tradeoff balanced toward exploitation when associative hierarchies are flat (as in the situation of collaborative search): in this situation, inconsistent tag choices for reoccurring topics

---

[1] We thank an anonymous reviewer for prompting this additional analysis.

may impair convergent group processes, i.e., reduce the benefits of exploiting each other's search results through a consistent tag vocabulary. We then compare SoMe to a most popular tags (MPT) recommender. SoMe's advantage over MPT should come to the fore under the collaborative search condition, that is, if the relationship between word and meaning (tag and topic) becomes looser (less consistent) due to flatter associative hierarchies.

In what follows, we first demonstrate this decreased semantic distinctiveness (looser tag topic relationship) in collaborative information search. This decreased semantic distinctiveness of popular tags should lead to problems when trying to predict and recommend tag choices by estimates of popularity, as in the case of MPT approaches.

In a second step, we then describe in detail how SoMe simulates a person's search of memory when choosing tags in order to filter an inconsistent tag vocabulary by searching for topically resonant (relevant) tags. Finally, we provide the results of the evaluation that has taken place in the same study reported above and addresses the second research question *whether SoMe achieves high tag acceptance rates under a collaborative search and, more specifically, whether the SoMe advantage (over MPT) is larger under the collaborative than individual search condition.*

3.1. The challenge of predicting tags based on flat associative hierarchies

Above, we have shown the shape of an associative hierarchy to affect tagging consistency, where a steep hierarchy gives rise to a closer relationship between tag and topic than a flat one (see Table 2). Furthermore, we argue that the extent of tagging consistency is related to the extent of semantic distinctiveness of popular tags (to be shown below), which in turn should determine the extent by which a TRM has to combine statistics of both popularity (tag use frequency) and semantic resonance (topical relevance of a tag) to derive appropriate recommendations.

For instance, in case of a highly consistent tagging behavior, a given tag w_i can be expected to be chosen quite exclusively for a certain topic t_j. More formally, if $P(t_j|w_i)$ denotes the posterior probability that w_i belongs to t_j, it should exhibit high estimates for only one of the eight search topics (of the current study) and rather low estimates for the remaining seven topics, given w_i is used consistently. The posterior probability can be calculated applying the Bayes theorem (e.g., Fu & Dong, 2012), where, in this study, the priors are estimated by counting their relative frequencies being assigned to bookmarks, and the conditional probability $P(w_i|t_j)$ by dividing n_j (number of bookmarks in t_j that are associated with w_i) by n (total number of bookmarks in t_j). To observe the shape of the average distribution of the posteriors over the eight topics (e.g., 'peaky' in case of consistency and flat in case of inconsistency), per tag we can – similar to a rank-frequency distribution – rank the eight topics according to their strengths, i.e., $P(w_i|t_j)$, then average these ranked posteriors across several tags (e.g., the seven most popular tags, MPT), and finally, as shown in the left diagram in Figure 6, draw these means against the corresponding ranks. In line with our expectation, the left diagram reveals that $P(w_i|t_j)$ drops off comparatively steeply under the individual search (solid line), which means that the steep associative hierarchies (reported above) are reflected in the emerging tag vocabulary by a steep hierarchy of the posteriors: given a particular topic, we can predict the choice of tags with higher certainty. By contrast, under the collaborative search, the flat associative hierarchies become manifest in a comparatively flat distribution of the posteriors (dashed line), indicating higher uncertainty when predicting tags based on a given set of resource topics.

As already argued, the strength of relationship between tag and topic should further affect the extent to which popular tags are topically distinct from each other. To quantify distinctiveness D_i of a tag w_i, we introduce the notion of a tag's semantic profile $\mathbf{S}_i$, which is simply the unranked distribution of $P(w_i|t_j)$ over the eight topics. D_i can then be defined as 1 minus the average cosine similarity between $\mathbf{S}_i$ and each of the
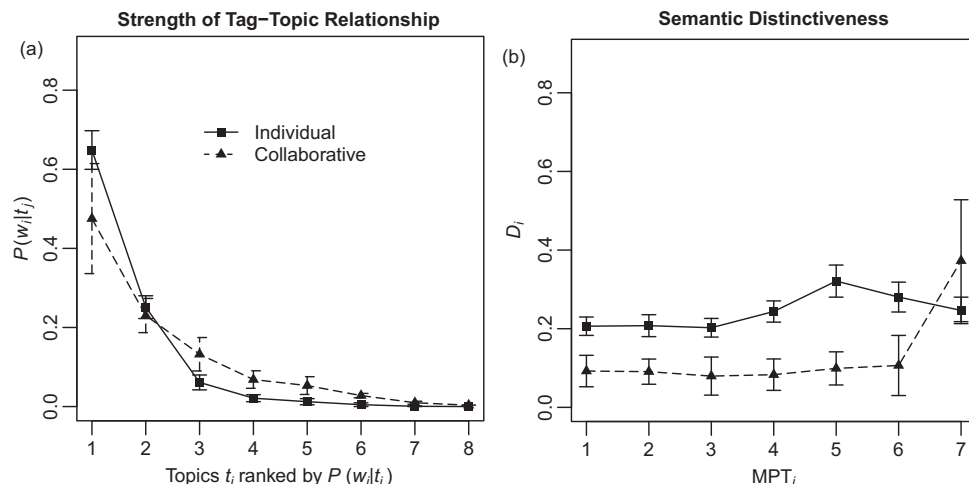

Figure 6. Strength of tag-topic relationship (left diagram, Figure 6a) and semantic distinctiveness (right diagram, Figure 6b) as a function of "Search Condition" (Individual vs. Collaborative). Error bars represent 1 standard error of mean.

profiles of the remaining six tags in MPT. In accordance with our expectation, the right diagram in Figure 6 shows D_i to be pronounced more strongly under the individual than collaborative search (except for the seventh tag in MPT), indicating a less ambiguous tag-topic co-occurrence pattern under the individual search, where different popular tags seem to refer to different search topics.

Summarizing the two patterns of Figure 6, we anticipate that the strategy of an MPT-based TRM, i.e., recommending tags from the head of the rank-frequency distribution, should be effective under the individual search condition: Due to its semantic distinctiveness and high popularity, the set of MPT (head of the rank-frequency distribution) is likely including a subset of familiar tags that fit the unknown resource an employee is bookmarking and tagging. Under the collaborative condition, however, statistics of popularity (rank and frequency) do not appear to provide sufficient information to disambiguate the tag recommendation problem: sampling from MPT does not necessarily result in a set of semantically distinct tags; instead it is likely to obtain a set of prominent tags indexing the same or similar topics. Thus, to identify tags that are both popular and topically distinct, a TRM strategy is needed filtering the tag vocabulary by popularity and semantic resonance (relevance).

3.2. Modeling search of memory to disambiguate the tag recommendation problem

To realize a TRM that searches for popular and semantically resonant (topically relevant) tags, we make use of a strategy that has been conceptually proposed (but not empirically evaluated) by Seitlinger, Ley, and Albert (2013). First simulation-based analyses of large-scale social tagging datasets (Kowald et al., 2014) have shown this strategy to be successful in modeling and predicting users' tag choices. The question, however, whether these results (i.e., high prediction accuracies) generalize to a realistic information search scenario, and whether tags can be recommended the employees actually adopt for their tag assignments, remains open and is addressed in the following.

This strategy implements the search of memory (SoMe) scheme (Figure 2), i.e., the cue-based activation of a search set S (step 1) and subsequent selection of N cue-relevant associations (tags; step 2). Drawing on MINERVA2, a model of episodic memory (e.g., Hintzman, 1986; Kwantes, 2005; Sprenger et al., 2011), the SoMe recommender distinguishes a primary memory (PM) that represents the experienced retrieval cue (e.g., the resource being bookmarked and tagged) from a secondary memory (SM) – "the vast pool of largely dormant memory traces" (Hintzman, 1986, p. 412). In the present study, the retrieval cue in PM is represented as the subset of search topics the employee assigns to the present resource via the bookmarking interface (Figure 4); a memory trace in SM is a record of each of the employee's bookmarks, in particular, of the correspondingly chosen topics and tags. The set of all SM traces therefore provides a memory-like representation of employee-specific topic-tag associations.

In the first step of the simulated PM-SM communication (cue-based activation of S), all traces in SM (topic-tag associations) are activated in parallel – according to their topical similarity to the cue (i.e., topic subset in PM): the more topics the cue and the trace have in common, the higher the trace's activation. That way, we implement a simple mechanism of semantic resonance, by which those tags come to the fore that are associated with cue-relevant topics. In the second step, the SoMe recommender performs a frequency-based ranking within the subset of strongly activated traces in order to select N cue-relevant associations (tags).

Summarizing, SoMe can be regarded a semantically resonant MPT approach that should improve MPT substantially, if semantic distinctiveness among MPT (see Figure 6) is low due to flat associative hierarchies. Based on the results of the first part that associative hierarchies are flatter and ideational fluency is higher under the collaborative than individual search, the third hypothesis H2 is: *The proposed tag recommender SoMe reaches higher levels of acceptance rate, i.e., is more likely recommending tags the user actually adopts for personal tag assignments, than MPT during a collaborative information search; no differences should be observed under the individual search condition.*

3.3. Method

This hypothesis (H2) was explored in the same study as previously reported.

Design

To answer the second research question and test hypothesis H2, the design (already described above) also included the independent variable 'Type of TRM' (MPT vs. SoMe; within subjects): When an employee bookmarked and tagged a new Web resource, (s)he could choose from a set of recommended tags that was generated either by MPT or SoMe – with an equal chance for either TRM being applied. Thus, to find an answer to our second research question, we realized a 2 (Search Condition: individual vs. collaborative search) × 2 (Type of TRM: MPT vs. SoMe) research design. The second dependent variable was acceptance rate of the tags generated by the two TRMs. Each of these variables and its operationalization are described below (see Section Measures). For the sake of a comprehensible presentation, however, we first describe some more details on the search task as well as the technical infrastructure.

Participants and search tasks

The process of annotating a bookmark, i.e., choosing topics and assigning tags, was supported by the bookmarking interface, illustrated by Figure 7, which this time also schematizes the tag recommenders' underlying algorithm (represented by the flow chart on the right-hand side): After selecting from the eight-topic list (number 2) to specify T (i.e., the subset of chosen topics) and clicking on the ok button, the algorithm was initiated producing a set of recommended tags, denoted W_{REC}: Depending on the outcome of the algorithm's first step, namely the random decision to choose between one of the two recommenders, W_{REC} was based on either the tags' frequency counts alone (in case of MPT) or on both the tags'
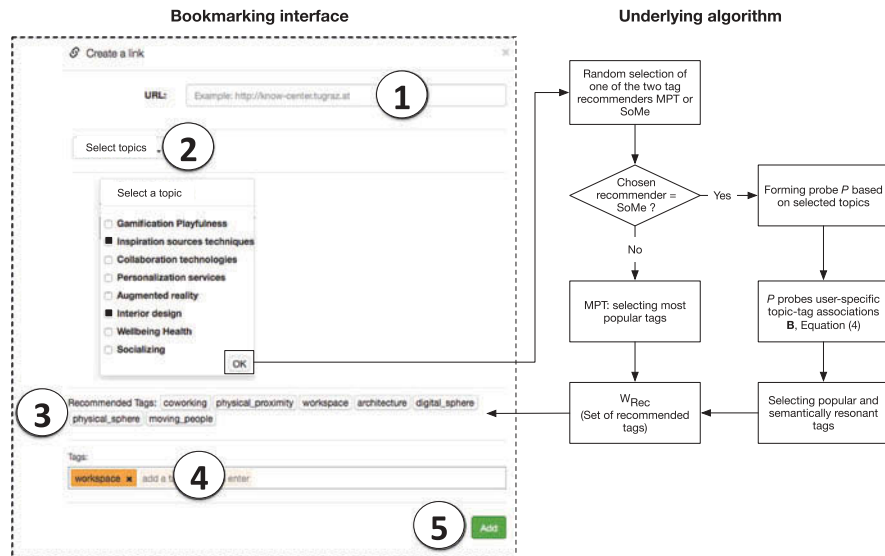
Figure 7. Bookmarking interface (left-hand side) and schematic illustration of the algorithm generating the set of recommended tags W_{REC} (right-hand side).

frequency counts and their semantic relatedness to T (in case of SoMe). The semantic relatedness was derived from a participant's past search behavior, in particular from topic-tag choices assigned to previously collected bookmarks. Details on how these choices were represented and processed by SoMe are given in the next section.

Finally, after displaying W_{REC} (number 3), the participant was free to choose from the recommended tag list as well as to type in personal tags (number 4). Note that this sequence of steps – encompassing the selection of topics, the algorithm, and the choice of tags – ensured that there were no interaction differences (from a user perspective) between the two recommenders, even if only SoMe was actually exploiting the topical information given by T.

Recommendation mechanisms

SoMe. The Search of Memory (SoMe) tag recommendation mechanism (TRM) was designed to mimic basic principles of an employee's search of memory when tagging a current Web resource. To this end, SoMe is based on the episodic memory model MINERVA2, which has been shown to account for a wide range of memory-based human behavior, such as recognition (Hintzman, 1988), categorization (Hintzman, 1986), representing word similarities (Kwantes, 2005) or judgments on the probabilities of hypotheses (Sprenger et al., 2011). Our goal, of course, was not to represent a user's entire episodic memory but to account for the encoding of episodic memory traces in the course of the user's ongoing search history. Beyond that, we did not aim to model specific search operations that would be formalized by more detailed models, such as CMR (e.g., Polyn, Norman, & Kahana, 2009), but to mimic the general principle of activating memory traces that are similar to a given environmental cue, i.e., the resource being tagged. To model this general principle, SoMe implemented the MINERVA2 distinction between a primary memory (PM) and secondary memory (SM) component. The role of PM was to represent the resource being tagged in terms of the chosen topics and to search SM for tags that the employee had assigned to

bookmarks with similar topics. To model PM, a probe P was formed as an 8-element, binary (1,0) feature vector – with each feature j representing one of the eight topics and taking on the value +1, if it was assigned to the current resource. SM was represented as a matrix **B**, with each row i being a binary (1,0) feature vector representing a bookmark in the employee's previous information search. The first $j = 1,...,8$ positions indexed the topic-features, the subsequent $j = (8 + 1),...,(8 + n)$ positions indexed the tag-features, where n was the number of tags generated within the whole bookmarking system.

To mimic a search of memory, SoMe proceeded in two steps: First, in a process of resonance, P interacted with the topic features ($j = 1,...,8$) of each row i in **B** to generate an overall value of activation $A(i)$ given by the cubed cosine between P and the row's first 8 feature values. Second, the mechanism estimated $R(j)$, i.e., the extent to which a feature resonated with the probe, by multiplying all features of each trace by the trace's activation $A(i)$ and then, summing these products across all m traces, given by

$$R(j) = \sum_{i=1}^{m} A(i)\mathbf{B}(i,j) \qquad (4)$$

Under the individual search condition, the tags, i.e., $j = (8 + 1),...,(8 + n)$, with the 7 highest resonance values $R(j)$ were included in W_{REC} and displayed in the bookmarking interface (number 3 in Figure 7). Under the collaborative condition, an average resonance value $\bar{R}(j)$ was calculated by mimicking a search of each employee's memory **B** and averaging across the individual $R(j)$ values. Finally, the seven highest ranked $\bar{R}(j)$ values were included in W_{REC}.

MPT. This type of TRM only considered the reuse count of tags and thus, determined the current rank-frequency distribution of all tags that had been generated in the shared (collaborative) or unshared (individual) bookmarking system. Then, the seven highest ranked tags were included into W_{REC} and displayed in the bookmarking interface.

Measures

Tag acceptance rate. The tag acceptance rate was measured calculating the precision and recall of the recommendations generated by MPT and SoMe. Each time an employee collected a new bookmark and classified it by selecting a set of topics, the bookmarking interface (Figure 7) displayed a set of $x = 7$ recommended tags (number 3 in Figure 7), denoted W_{REC}, and the employee was free to type in personal tags and/or to choose tags from W_{REC}. If W_{APP} denotes the set of tags actually applied by the employee, the metrics of *precision* and *recall* are calculated by the following two fractions:

$$precision = |W_{APP} \cap W_{REC}|/|W_{REC}| \quad (5)$$

$$recall = |W_{APP} \cap W_{REC}|/|W_{APP}| \quad (6)$$

To combine both metrics and to derive a single score, the harmonic mean of *precision* and *recall* is calculated, which is denoted *F*-score and given by

$$F = (precision * recall * 2)/(precision + recall) \quad (7)$$

Note that these metrics depend on the number of elements (tags) included in W_{REC}, and thus, precision, recall and the *F*-score are usually determined for each possible *x*, which varied between $x = 1$ and $x = 7$ in the present study. Following this evaluation practice, we determined the average *F*-score under the four factorial combinations (of the 2 × 2 research design) and for every *x*. For statistical analysis, we then performed a 2 (Search Condition) × 2 (Type of TRM) repeated measures ANOVA on the *F*-score, taking *x* as the unit of analysis.

3.4. Results

We expected an interaction of the search condition with the performance of the tag recommender SoMe, i.e., a larger advantage over the baseline recommender MPT under the collaborative than individual condition (H2).

Hypothesis H2

The tag recommender SoMe was designed to improve the recommendation of tags under a collaborative condition, i.e., conditions of flat associative hierarchies and low semantic distinctiveness. We, therefore, hypothesized an interaction between 'Type of TRM' (MPT vs. SoMe) and 'Search Condition' with respect to the tag acceptance rate and in particular, a larger SoMe advantage under the collaborative condition. To test this interaction, we extracted all tagging events under each of the four factorial combinations and compared the set of recommended tags (W_{REC}) with the set of actually applied tags (W_{APP}) by calculating the *F*-score (Equation 7). Figure 8 presents the results. Under the individual condition (left diagram), the two recommenders appear to reach similar estimates of the *F*-score for a varying number *x* of recommended tags (drawn on the abscissa). Averaged over *x*, the *F*-scores achieved by MPT and SoMe are 0.29 ($SD = 0.04$) and 0.30 ($SD = 0.03$), respectively. We therefore conclude that during an individual information search, where employees exhibit comparatively steep associative hierarchies, both MPT and SoMe generated recommendations at a comparable acceptance rate.

As expected, the relation between the two recommenders changed under the collaborative condition (right plot), where, descriptively, SoMe reached higher estimates of the *F*-score than MPT across all values of *x*. In this case, the average and *x*-independent scores for SoMe and MPT are 0.34 ($SD = 0.09$) and 0.27 ($SD = 0.08$), respectively. Thus, we observed an interaction between the variables 'Search Condition' and 'Type of TRM' and found that SoMe outperformed MPT only during a collaborative information search. The results of the ANOVA supported this interpretation by yielding a significant effect for the interaction of the two variables, $F(1,6) = 12.45$, $p < .05$. Beyond that, the test yielded a significant main effect for 'Type of TRM', $F(1,6) = 53.96$, $p < .001$, but not for 'Search Condition', $F(1,6) = 0.19$, *n.s.* Due to the interaction, we did not consider the main effect and concluded in line with hypothesis H2 that the SoMe-advantage over MPT only applies to the collaborative condition.
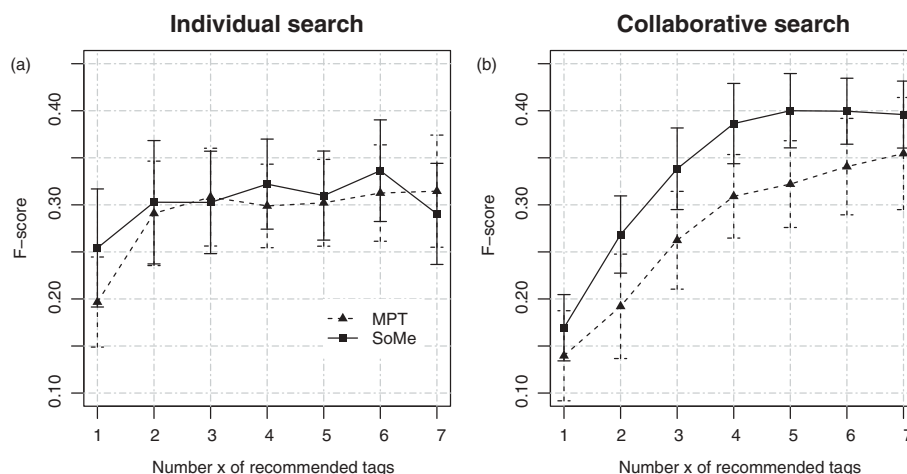

Figure 8. Tag acceptance rate achieved by the MPT and SoMe for a varying number of recommended tags (*x* = 1,...,7) under the individual (left, Figure 8a) and collaborative search (right, Figure 8b). Error bars represent 1 standard error of mean.

3.5. Discussion

These results demonstrate that the MPT-based approach is less appropriate to predict and support individual tagging behavior, if it is embedded into a collaborative search scenario where associative hierarchies tend to be flat. A considerably more effective strategy to derive topically relevant recommendations under inconsistent tagging patterns appears to be the simulation of a resource-triggered search through associative memory for familiar and semantically resonant words.

Therefore, the proposed recommender SoMe seems to implement a promising and robust approach to increase employees' tendencies of reusing existent tags independent of how flat or steep the underlying associative hierarchies are. Hence, it appears to be an effective and expandable strategy for a non-human actor (i.e., TRM) helping to balance exploration/divergence and exploitation/convergence during a networked search. This is particularly the case in a collaborative situation, where ideational fluency is high (driving individual exploration and experimentation) and tagging consistency needs to be supported to also drive exploitation of each other's search results. In such situation of higher inconsistency, a given topic is more likely to co-occur with several tags with similar frequencies. As a consequence, the rank-frequency distribution of the tag vocabulary becomes flatter and popular tags (head of the rank-frequency distribution) start overlapping in the topics they are used for. Thus, statistics of popularity (e.g., usage frequency) alone no longer suffice to identify topically distinct popular tags, among which a subset would be likely to match a current resource's topics. In contrast, SoMe overcomes this problem of decreased distinctiveness by filtering the tag vocabulary not only by popularity but also semantic resonance.

By doing so, SoMe considers more information about an employee's tagging behavior than MPT does, and therefore a general advantage of SoMe over MPT (a "main effect") is of course expected. However, the question we have explored with our study is not whether SoMe in general is a more effective recommender than MPT, but rather whether the advantage of SoMe interacts with the Search Condition, i.e., whether the advantage is larger in the collaborative search condition. We are also not suggesting SoMe as the most effective recommendation strategy in collaborative information search. To be able to claim this would require a broader set of alternative strategies to compare SoMe to. In this article, we were rather interested in examining some of the effects of collaborative information search on more fundamental cognitive processes such as memory retrieval.

4. Overall discussion

The first goal of the study was to investigate the question whether mutual stimulation during a collaborative and Web-based information search has an impact on an employee's mental organization of associations around a given topic. In particular, we expected a tag-mediated circulation of ideas (collaborative search) to let an employee's associative hierarchy become flatter, which should be indicated by a tradeoff between an increase in ideational fluency and a decrease in tagging consistency. The second goal was to introduce and test the tag recommendation mechanism SoMe designed to push the reuse of tags and thus, to compensate for the hypothesized inconsistency in a collaborative search that benefits from balancing processes of divergence (e.g., fluency) and convergence (e.g., consistency; Lazer & Bernstein, 2012).

First, the task of searching for Web resources was found to yield a general learning effect independent of the Search Condition (individual vs. collaborative) by extending the associative structure by topic-relevant representations. Additionally, as this increase of associations did not reduce their availability (response time), we concluded that learning consisted not only in broadening the fan of topic-relevant associations (i.e., flattening the associative hierarchy) but also in inhibiting or even excluding topic-irrelevant associations. The resulting associative structure gave rise to an increased ideational fluency, which was characterized by both a larger number and a higher speed of responses. Beyond that, this effect appeared to be larger under the collaborative than individual condition and became manifest in a fluency-consistency tradeoff. In line with hypothesis H1.1., the flatter and more topic-related hierarchy under the collaborative condition allowed for a steadier stream of ideas in response to a particular topic (e.g., 'interior design'); at the same time, it also caused a more variable assignment of tags to re-occurring topics during Web search, thereby lending support to H1.2.

From a psycho-pedagogical perspective, we therefore gained evidence of a positive impact of mutual stimulation on individual learning processes: a higher ideational fluency facilitates creative cognition to the extent that it increases the probability of bringing otherwise dissociated ideas for a new and useful combination into ideational contiguity (e.g., Benedek & Neubauer, 2013; Mednick, 1962). Also from an information discovery (Kerne et al., 2008) and information-based ideation viewpoint (Kerne et al., 2014), our results have practical design implications by revealing that an individual is more likely to experience cognitive restructuring and to forge new associations when searching the Web, if she or he participates in a collaborative and tag-based curation of an evolving problem space. From the perspective of information retrieval, however, a tradeoff on the expense of consistency exacerbates the stabilization of the tagging vocabulary and thus, organization of an already curated space.

In order to compensate for this downside of mutual stimulation and to support a tag-based collaborative information search, we deduced and evaluated the tag recommendation mechanism (TRM) SoMe from the search of memory scheme. By mimicking a resource-triggered search of associative memory (topic-tag associations) to identify and recommend topically relevant tags, SoMe achieved adequate tag acceptance rates and, under the collaborative condition, outperformed a baseline TRM built upon the most popular tags (MPT) approach (hypothesis H2). The SoMe advantage during a collaborative search provided evidence of the model's validity (in representing an associative hierarchy) and showed an increased availability of a larger number of associations to be mirrored by a looser tag-topic relationship. Under such conditions of inconsistency, a tag's topical relevance (modeled

by the search through memory) had to be taken into account in addition to the tag's popularity (modeled by frequency) in order to anticipate an employee's tagging behavior accurately. Again our model of reflective search would predict such interaction: Under the individual search condition, TRM and tag clouds create a self-reinforcing loop between the individual and her or his own artifacts that strengthens tag associations to only a few and recurrently reflected search topics. The consequence is a steep associative hierarchy with semantically distinct tags ranking at the hierarchy's top. Under the collaborative condition, however, TRM and tag clouds take a propagating role and animate an individual to reflect and experience tags in a drifting context of varying topics. The consequences are broader and overlapping fans of topic associations that are hardly distinguishable based on popularity.

5. Conclusions and practical implications

The study reported here adds further evidence to the reflective search framework that we have developed in the context of social tagging on the web. Through this study, we have added evidence through two diverse research methods. First and in a more traditional way, we test the effects of information search on mental organization (associative hierarchy). By modeling and simulating memory access mechanisms with a TRM, we then gather further evidence for the soundness of the approach by means of an entirely new research strategy. Taken together, we believe these results offer convergent evidence for the framework we have been proposing.

This approach is similar to that of Nijstad and Stroebe (2006) in their research on group creativity. Like these authors, we show how a model of learning and search of memory can become part of a socio-cognitive framework (i.e., reflective search) and how this integration helps to describe and simulate supra-individual processes of tag-based mutual stimulation. In our understanding, social tags that people encounter in a shared environment do not lead to imitation, but rather trigger a reflection of the search topic which changes their mental structures. The model of search through memory (see Figure 2) serves as starting point to specify the notion of an associative hierarchy that we assume to be enacted when users reflect on objects (e.g., problems, Web resources; Seitlinger & Ley, 2016). The evidence we found for changes in that associative hierarchy lends further evidence to the assumption that reflection triggers an individual learning process (e.g., Renner, Prilla, Cress, & Kimmerle, 2016). This act of "learning-by-reflecting" is stimulated by joint artifacts (e.g., social tags) and thus, contributing to a collaborative search results in a broader fan of topic-related associations and consequently, an increased level of ideational fluency.

As ideational fluency is an indicator of creative potential (e.g., Benedek & Neubauer, 2013), the present study also sheds some light on a process that appears to be conducive to group creativity (e.g., Nijstad & Stroebe, 2006; Paulus & Brown, 2007), namely artifact-mediated learning-by-

reflecting, and how this process can be facilitated by means of creativity-support tools (e.g., Kerne et al., 2008; Sarmiento & Stahl, 2008), namely tagging combined with an effective tag recommendation mechanism (TRM). Especially in the light of echo chamber effects (Pariser, 2011; Sunstein, 2001), which are frequently observed in online discourses, intelligent information services become more important that prevent a group from converging in a particular perspective too early. In future work, we therefore would like to investigate the question whether the applied recommender approach of providing stimuli that are novel but also resonate with personal knowledge structures can be an effective way to maintain a distribution of diverse ideas and thus, to counteract echo chamber effects.

The present work also has implications for the application of memory models in the context of TRM. While in the present work, an abstract scheme of memory search and a simple computational interpretation sufficed to account for a comparatively simple behavior (i.e., choosing tags), continuative research questions on Human-Web interactions could require more complex models of memory search. The present study took into account two landmarks of memory psychology in recommending tags, namely *familiarity* (represented by the popularity of tags), and *semantic resonance* (as specified by the topical filtering through the MINERVA model). In the future, one goal will be to further refine the SoMe strategy, for instance, by extending the model by *recency* of tag use (e.g., Trattner, Kowald, Seitlinger, Kopeinik, & Ley, 2016), another important factor in retrieval from memory. Furthermore, context retrieval models (e.g., Polyn et al., 2009) would allow specifying and parameterizing executive processes (e.g., internal context updates) that take part in higher-level cognition (e.g., reflection and formation of information goals) during Web exploration (Seitlinger & Ley, 2016).

6. Limitations and future work

One limitation of this study is the small sample size of $N = 18$ participants, which is owed to the difficulty of acquiring employees motivated to take part in a four-weeks work-integrated field study. To face this problem, we took particular statistical measures inspired by previous studies on social tagging, such as the one of Pirolli and Kairam (2012) or the one of Fu and Dong (2012) who performed model-based analyses of data from samples of $N = 18$ and $N = 8$ participants, respectively. For instance, to investigate hypothesis H1.1, we drew on a well-established modeling technique of memory psychology, which allowed deriving robust estimates (of N and λ) by aggregating responses across both different stimuli (i.e., topics) and different participants. Furthermore, to examine H1.2, not the individual participant but her or his recorded activities within the bookmarking system served as units of analysis, resulting in 685 rather than 18 data points. Finally, for analyzing H2, the unit of analysis was x (i.e., the number of recommended tags), which again allowed for aggregating across participants and all their

bookmarks and hence, deriving robust estimates of their tag acceptance rates.

While these statistical techniques help mitigate the problem of small sample sizes, they do not dissolve it. Therefore, future studies are necessary that clarify whether the reported results can be replicated and whether they generalize to different conditions. On the one hand, such studies will have to explore the impact of participants' background and level of expertise in searching for information. The present sample consisted of knowledge workers who can be assumed to be highly trained in performing the activities addressed by our research questions. Thus, even if we assume general-psychological processes are at play, such as cue-dependent search of memory, which should generalize across knowledge domains and skill levels, questions around external validity remain and must be clarified empirically. On the other hand, generalizability will have to be checked in terms of software features that create affordance for particular behavior. For instance, the eight-topic list of the bookmarking interface could have prompted specific priming processes when participants assigned tags to their collected resources and thus, affected their tagging consistency. Though statistical control analyses indicated no substantial impact arising from the number of chosen topics, future experiments are necessary to systematically observe the effect of a more or less differentiated upper-level structure.

Disclosure of potential conflicts of interest

No potential conflict of interest was reported by the author.

Funding

References

Albert, D. (1968). Free recall of word sequences as stochastic storage removal. *Zeitschrift Für Experimentelle Und Angewandte Psychologie, 15,* 564–581.

Benedek, M., & Neubauer, A. (2013). Revisiting Mednick's model on creativity-related differences in associative hierarchies. Evidence for a common path to uncommon thought. *The Journal of Creative Behavior, 47*(4), 273–289. doi:10.1002/jocb.35

Bousfield, W. A., & Sedgewick, C. H. W. (1944). An analysis of restricted associative responses. *Journal of General Psychology, 30*(2), 149–165. doi:10.1080/00221309.1944.10544467

Bousfield, W. A., Sedgewick, C. H. W., & Cohen, B. W. (1954). Certain temporal characteristics of the recall of verbal associates. *American Journal of Psychology, 67*(1), 111–118. doi:10.2307/1418075

Candy, L., & Hewett, T. (2008). Special issue introduction: Investigating and cultivating creativity. *International Journal of Human-Computer Interaction, 24*(5), 441–443. doi:10.1080/10447310802142060

Davelaar, E. J., & Raaijmakers, J. G. W. (2012). Human memory search. In P. M. Todd, T. T. Hills, & T. W. Robbis (Eds.), *Cognitive search* (pp. 177–194). Cambridge, MA: MIT Press.

Dellschaft, K., & Staab, S. (2012). Measuring the influence of tag recommender on he indexing quality in tagging systems. In E. Munson & M. Strohmaier (Eds.), *Proceedings of the 23rd ACM conference on hypertext and social media* (pp. 73–82). New York, NY: ACM Press. doi:10.1145/2309996.2310009

Dorfman, L., Martindale, C., Gassimova, V., & Vartanian, O. (2008). Creativity and speed of information processing: A double dissociation involving elementary versus inhibitory cognitive tasks. *Personality and Individual Differences, 44*(6), 1382–1390. doi:10.1016/j.paid.2007.12.006

Font, F., Serrà, J., & Serra, X. (2015). Analysis of the impact of a tag recommendation system in a real-world folksonomy. *ACM Transactions on Intelligent Systems and Technology, 7*(1). doi:10.1145/2743026

Fu, W.-T., & Dong, W. (2012). Collaborative indexing and knowledge exploration: A social learning model. *IEEE Intelligent Systems, 27*(1), 39–46. doi:10.1109/MIS.2010.131

Fu, W.-T., Kannampallil, T., Kang, R., & He, J. (2010). Semantic imitation in social tagging. *ACM Transactions on Computer-Human Interaction, 17*(3). doi:10.1145/1806923.1806926

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communications of the ACM, 30*(11), 964–971. doi:10.1145/32206.32212

Golder, S. A., & Huberman, B. A. (2006). The structure of collaborative tagging systems. *Journal of Information Science, 32*(2), 198–208. doi:10.1177/0165551506062337

Hintzman, D. L. (1986). "Schema abstraction" in a multiple trace-model. *Psychological Review, 93*(4), 411–428. doi:10.1037/0033-295X.93.4.411

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551. doi:10.1037/0033-295X.95.4.528

Hutchins, E., & Johnson, C. (2009). Modeling the emergence of language as an embodied collective cognitive activity. *Trends in Cognitive Science, 1*(3), 523–546. doi:10.1111/j.1756-8765.2009.01033.x

Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes.* Boston, MA: Houghton Mifflin.

Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., & Stumme, G. (2007). Tag recommendations in folksonomies. In *Knowledge discovery in databases: PKDD 2007. Lecture notes in computer science* (pp. 506–514). Berlin, Heidelberg: Springer. doi:10.1007/978-3-540-74976-9_52

Kaplan, I. T., Carvellas, T., & Metlay, W. (1969). Searching for words in letter sets of varying sizes. *Journal of Experimental Psychology, 82*(2), 377–380. doi:10.1037/h0028140

Kerne, A., Smith, S., Koh, E., Choi, H., & Graeber, R. (2008). An experimental method for measuring the emergence of new ideas in information discovery. *International Journal of Human-Computer Interaction, 24*(5), 460–477. doi:10.1080/10447310802142243

Kerne, A., Webb, A., Smith, S., Linder, R., Lupfer, N., Qu, Y., … Damaraju, S. (2014). Using metrics of curation to evaluate information-based ideation. *ACM Transactions on Computer-Human Interaction, 21*(3), 14:1-14:48. doi:10.1145/2591677

Kohn, N., Paulus, P., & Choi, Y. (2011). Building on the ideas of others: An examination of the idea combination process. *Journal of Experimental Social Psychology, 47,* 554–561. doi:10.1016/j.jesp.2011.01.004

Kowald, D., Seitlinger, P., Kopeinik, S., Ley, T., Albert, D., & Trattner, C. (2014). Forgetting the words but remembering the meaning: Modeling forgetting in a verbal and semantic tag recommender. In M. Atzmueller, A. Chin, C. Scholz, & C. Trattner (Eds.), *Mining, modeling, and recommending things in social media* (pp. 75–95). Heidelberg, Germany: Springer Lecture Notes in Artificial Intelligence. doi:10.1007/978-3-319-14723-9_5

Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin & Review, 12*(4), 703–710. doi:10.3758/BF03196761

Latour, B. (2005). *Reassembling the social. An introduction to actor-network theory.* Oxford, New York: Oxford University Press.

Lazer, D., & Bernstein, E. S. (2012). Problem solving and search in networks. In P. M. Todd, T. T. Hills, & T. W. Robbis (Eds.), *Cognitive search* (pp. 269–282). Cambridge, MA: MIT Press.

Lazer, D., & Friedman, A. (2007). The network structure of exploration and exploitation. *Administrative Science Quarterly*, 52, 667–694. doi:10.2189/asqu.52.4.667

Linder, R., Snodgrass, C., & Kerne, A. (2014). Everyday ideation: Al of my ideas are on interest. In A. Schmidt & T. Grossman (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2411–2420). New York, NY: ACM press. doi:10.1145/2556288.2557273

Lorince, J., & Todd, P. (2016). Music tagging and listening: Testing the memory cue hypothesis in a collaborative tagging system. In M. Jones (Ed.), *Big data in cognitive science*. New York, NY: Psychology Press.

Martindale, C. (1995). Creativity and connectionism. In S. Smith, T. Ward, & R. Finke (Eds.), *The creative cognition approach* (pp. 249–268). Cambridge, MA: MIT Press.

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review*, 69(3), 220–232. doi:10.1037/h0048850

Nelson, L., Held, C., Pirolli, P., Hong, L., Schiano, D., & Chi, E. H. (2009). With a little help from my friends: Examining the impact of social annotations in sensemaking tasks. In K. Hinckley, M. Ringel, S. Hudson, & S. Greenberg (Eds.), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1795–1798). New York, NY: ACM press. doi:10.1145/1518701.1518977

Nijstad, B., & Stroebe, W. (2006). How the group affects the mind: A cognitive model of idea generation in groups. *Personality and Social Psychology Review*, 10(3), 186–213. doi:10.1207/s15327957pspr1003_1

Page, S. E. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press.

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. New York, NY: Penguin Press.

Paulus, P., & Brown, V. (2007). Towards more creative and innovative group idea generation: A cognitive-social-motivational perspective of brainstorming. *Social and Personality Psychology Compass*, 1(1), 248–265. doi:10.1111/j.1751-9004.2007.00006.x

Pirolli, P., & Kairam, S. (2012). A knowledge-tracing model of learning from a social tagging system. *User Modeling and User-Adapted Interaction*, 23(2), 139–168. doi:10.1007/s11257-012-9132-1

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116(1), 129–156. doi:10.1037/a0014420

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol 14, pp. 207–262). New York, NY: Academic Press.

Renner, B., Prilla, M., Cress, U., & Kimmerle, J. (2016). Effects of prompting in reflective learning tools: findings from experimental field, lab, and online studies. *Frontiers in Psychology*, 7. http://doi.org/10.3389/fpsyg.2016.00820

Roepstorff, A., Niewöhner, J., & Beck, S. (2010). Enculturing brains through patterned practices. *Neural Networks*, 23(8–9), 1051–1059. doi:10.1016/j.neunet.2010.08.002

Rohrer, D. (2002). The breadth of memory search. *Memory*, 10(4), 291–301. doi:10.1080/09658210143000407

Sarmiento, J. W., & Stahl, G. (2008). Group creativity in interaction: Collaborative referencing, remembering, and bridging. *International Journal of Human-Computer Interaction*, 24(5), 492–504. doi:10.1080/10447310802142300

Schweiger, S., Oeberst, A., & Cress, U. (2014). Confirmation bias in web-based search: A randomized online study on the effects of expert information and social tags on information search and evaluation. *Journal of Medical Internet Research*, 16(3), e94. doi:10.2196/jmir.3044

Seitlinger, P., & Ley, T. (2012). Implicit imitation in social tagging: Familiarity and semantic reconstruction. In E. H. Chi & K. Höök (eds), *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1631–1640). New York, NY: ACM press. doi:10.1145/2207676.2208287

Seitlinger, P., & Ley, T. (2016). Reconceptualizing imitation in social tagging: A reflective search model of human web interaction. In P. Parigi & S. Staab (Eds.), *Proceedings of the 8th ACM conference on web science* (pp. 146–155). New York, NY: ACM press. doi:10.1145/2908131.2908157

Seitlinger, P., Ley, T., & Albert, D. (2013). An implicit-semantic tag recommendation mechanism for socio-cognitive learning systems. In T. Ley, M. Ruohonen, M. Laanpere, & A. Tatnall (Eds.), *Proceedings of Open and Social Technologies for networked learning OST'12* (pp. 41–46). Heidelberg, Germany Springer LNCS. doi:10.1007/978-3-642-37285-8_5

Seitlinger, P., Ley, T., & Albert, D. (2015). Verbatim and semantic imitation in indexing resources on the Web: A fuzzy-trace account of social tagging. *Applied Cognitive Psychology*, 29(1), 32–48. doi:10.1002/acp.3067

Smith, S. M., Ward, T. B., & Finke, R. A. (1995). *The creative cognition approach*. Cambridge, MA: MIT Press.

Sprenger, A., Dougherty, M., Atkins, S., Franco-Watkins, A., Thomas, R., Lange, N., & Abbs, B. (2011). Implications of cognitive load for hypothesis generation and probability judgment. *Frontiers in Psychology*, 2. doi:10.3389/fpsyg.2011.00129

Sunstein, C. (2001). *Echo chambers: Bush v. Gore, Impeachment, and Beyond*. Princeton, Oxford: Princeton University Press.

Trattner, C., Kowald, D., Seitlinger, P., Kopeinik, S., & Ley, T. (2016). Modeling activation processes in human memory to predict the use of tags in social bookmarking systems. *The Journal of Webscience*, 2(1), 1–16. doi:10.1561/106.00000004

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114(1), 104–132. doi:10.1037/0033-295X.114.1.104

Wagner, C., Singer, P., Strohmaier, M., & Huberman, B. (2014). Semantic stability in social tagging streams. In A. Broder, K. Shim, & T. Suel (Eds.), *Proceedings of the 23rd international conference on world wide web* (pp. 735–746). New York, NY: ACM Press. doi:10.1145/2566486.2567979

Ward, T. B. (2007). Creative cognition as a window on creativity. *Methods*, 42(1), 28–37. doi:10.1016/j.ymeth.2006.12.002

Webster, J., Gibbins, N., Halford, S., & Hracs, B. (2016). Towards a theoretical approach for analyzing music recommender systems as sociotechnical cultural intermediaries. In P. Parigi & S. Staab (Eds.), *Proceedings of the 8th ACM Conference on Web Science* (pp. 137–145). New York, NY: ACM press. doi:10.1145/2908131.2908148

Wixted, J., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review*, 1(1), 10–89. doi:10.3758/BF03200763

About the Authors

Dr Paul Seitlinger is a senior researcher at the School of Educational Sciences (Tallinn University). His main interest is the role of human memory in reflective search on the Web and in the design of ideationally stimulating information services. Paul has a PhD in Psychology from the University of Graz.

Dr Tobias Ley is Professor of Learning Analytics and Educational Innovation (Tallinn University). He has published over 100 publications in leading outlets in HCI and technology-enhanced learning. He has had leading roles in a number of large-scale EU-funded projects. Tobias has a PhD in Psychology from the University of Graz.

Dr Dominik Kowald is university assistant at the Institute of Interactive Systems and Data Science (Graz University of Technology). Additionally, he is senior researcher in the Social Computing research area of Know-Center Graz. He has recently finished his PHD thesis on cognitive-inspired recommender systems for social tagging environments

Dipl.-Ing. Dieter Theiler is software engineer at Know-Center Graz (Social Computing research area). He studied Software Engineering and Economy at Graz University of Technology and was responsible for the software engineering process in the course of the EU-IP Learning Layers. His interests include software architecture and engineering, TEL and CSCW.

Dipl.-Ing. Ilire Hasani-Mavriqi is PhD student at the Institute of Interactive Systems and Data Science (Graz University of Technology) and part of the Social Computing team at Know-Center Graz. Her research interests include social network analysis, dynamics in online networks, opinion diffusion and dynamics of social influence in complex networks.

Mag. *Sebastian Dennerlein* is deputy head of the Social Computing research area and business developer at Know-Center Graz. He studied Psychology at the University of Graz and is doing his PhD on Collaboration and Meaning Making in socio-technical systems. His research interests are cognitive psychology, TEL, CSCL, and CSCW.

Dr Elisabeth Lex, Asst.-Professor at Graz University of Technology, heads the Social Computing research area at Know-Center Graz. Her research interests include Dynamics in Complex Networks, Recommender Systems, Web Science, Open Science and Machine Learning. She is also member of the Expert Group on Altmetrics, which advises the European Commission.

Dr Dietrich ALBERT, Univ.-Professor, is (a) Senior Scientist at the *Graz University of Technology* (b) Prof. em. at the *University of Graz*, and (c) Key Researcher at the *Know-Center*. Combining *Cognitive Science* and *Information Communication Technology* in different fields is the focus of his current research, development, and innovation (http://css-kmi.tugraz.at/research/projectvis/tagcloud.html).