Full length article

For they know not what they do? Target memory and metacognitive monitoring of self-disclosures on social networking sites

Stephanie Pieschl[*], Ricarda Moll [1]

Westfälische Wilhelms-Universität, Institut für Psychologie, Fliednerstr. 21, 48149 Münster, Germany

A B S T R A C T

To adequately weigh the risks and benefits of self-disclosures on Social Networking Sites (SNS), it is pivotal for users to be aware of the current status of their digital privacy. Cognitive and metacognitive factors seem necessary in this context, for example *target memory* about who exactly has access to exactly which pieces of information and *metacognitive monitoring* of the status of one's own target memory. Transferring paradigms from basic memory and metacognition research, we systematically investigated the impact of *risk cues* on these variables: In Experiment 1 ninety-eight young SNS users repeatedly disclosed personal or impersonal information (between-subject: information intimacy) to large or small audiences (within-subject: audience size). Afterwards their target memory was assessed and they gave metacognitive confidence judgments regarding each answer. Results indicate that participants remembered well if they had disclosed something but struggled with target memory (correct *association between content and specific audience*), an effect that was mirrored regarding metacognitive monitoring. Importantly participants' target memory was significantly better under conditions of *risk* (personal information, large audience), but they were not metacognitively aware of these risk effects. Experiment 2 ($N = 59$) replicated the audience size effects and confirmed that these were not artifacts of visual salience.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction to digital privacy regulation

Most people in developed countries have access to the Internet and use it frequently via computers, laptops, tablets, smart phones, and other technical devices. Social Media, especially Social Networking Sites (SNS), are one of the most frequently used applications on the Internet (Feierabend, Plankenhorn, & Rathgeb, 2015). One of the key features of SNS is that users disclose information about themselves. Therefore, competent digital privacy regulation is one of users' greatest challenges when using Social Media (Trepte & Dienlin, 2014). Unlike in Face-to-Face (FtF) contexts, users' posted digital data is persistent, searchable, scalable and replicable (boyd, 2008). Therefore, specific risks are associated with digital self-disclosures, for example data access by unauthorized people (Attrill & Jalil, 2011).

Digital privacy regulation can be conceptualized as a specific example of self-regulation. Self-regulation underlies all goal-directed behavior (Miller, Galanter, & Pribram, 1960): Organisms set desired *target* states, diagnose their *current* states, and regulate their environment, behavior, or their internal processes to reduce potential *discrepancies* between these two. Metacognitions about one's own cognitions (Flavell, 1979) are at the core of such self-regulatory processes: For example, processes of metacognitive *monitoring* are necessary to diagnose one's current state of learning, and processes of metacognitive *control* are necessary to initiate, for example, the use of more effective learning strategies (Nelson & Narens, 1994). Social Media users are assumed to self-regulate their digital privacy similarly: They set a target state of digital privacy, diagnose their current state of digital privacy, and actively regulate their digital privacy if they detect discrepancies, for example by additional self-disclosures (Petronio, 2002; Trepte & Dienlin, 2014).

We proposed that basic cognitive and metacognitive processes should be highly relevant for digital privacy regulation in SNS (Moll, Pieschl, & Bromme, 2014b). Similar to so-called target memory in

* Corresponding author.
  E-mail addresses: pieschl@uni-muenster.de (S. Pieschl), ricarda.moll@uni-muenster.de (R. Moll).
  [1] Ricarda Moll is now at Verbraucherzentrale NRW e.V. (National Consumer Centre, North-Rhine Westphalia, registered association), Mintropstr. 27, 402015 Düsseldorf, Germany.

FtF contexts, users need to accurately remember what they have disclosed in the past and who exactly has access to that information in order to have an accurate representation of their *current* state of digital privacy. Otherwise users might be mistakenly convinced that only little information about them is available on the Internet and might ultimately mistake the de facto cumulative risk of digital self-disclosing as a situational one (McCloy, Byrne, & Johnson-Laird, 2010; Slovic, 2000). Additionally, if users became aware of their own target memory problems in a SNS context, they might be able to compensate for these problems by deliberately looking up their previous digital self-disclosures, for example their Facebook profiles and corresponding privacy settings. Thus, they could improve their diagnosis of their current state of digital privacy. Therefore, accurate metacognitive monitoring of one's own target memory should also play a pivotal role in Social Media users' digital privacy regulation. In the subsequent sections we elaborate theoretically why and how target memory and metacognitive monitoring should be relevant to digital privacy regulation and how risk could alleviate potential problems.

2. Target memory

During FtF communication people have to track both, the sources of incoming information as well as the targets of outgoing information. Regarding targets, they need to remember the communication topics and communication partners separately, but even more importantly they need to remember the correct content-target associations, namely *which detail was told to whom* (so-called target memory). Otherwise, people might repeatedly tell the same stories to the same people or simply misjudge common ground; in other words, target memory facilitates everyday communication (Gopie & MacLeod, 2009). Up to now, target memory has exclusively been studied in FtF contexts. In a prototypical paradigm participants are placed in fictitious communication situations where they have to tell multiple facts to multiple persons; afterwards their target memory is assessed and, for example, compared to their memory for facts alone (El Haj, Postal, & Allain, 2013; Gopie & MacLeod, 2009; Gopie, Craik, & Hasher, 2010). Results show that people generally remember their communication partners and conversation topics quite well separately, but struggle with target memory (Brown, Hornstein, & Memon, 2006; Gopie & MacLeod, 2009; Gopie et al., 2010; Marsh & Hicks, 2002).

Some findings shed further light onto cognitive mechanisms underlying target memory: In most studies, source memory was superior to target memory (Gopie & MacLeod, 2009; Koriat, Ben-Zur, & Druch, 1991), but this pattern reversed when participants could actively decide about their targets whenever they revealed information (Marsh & Hicks, 2002). Furthermore, target memory improved when generative activities such as active self-disclosures were involved in revealing information compared to read-only or imagine-only conditions (El Haj et al., 2013; Koriat et al., 1991). Additionally, target memory improved when participants focused on the targets (by saying names of communication partners) and declined when they focused on themselves (by telling personal facts) (Gopie & MacLeod, 2009). Thus, *active decisions*, *generative activities*, and *focusing on the target context* during encoding improved later target memory.

Communication on SNS, however, is inherently *decontextualized* with little information about potential targets and SNS communication often involves *few active decisions*. For example, target cues such as facial expressions are non-existent (Kiesler, Siegel, & McGuire, 1984) and other contextual cues are constant in all Social Media communication episodes within the same application. Active decisions regarding target audiences are also unlikely to occur, as the majority of users alter their privacy settings only once

when they join their SNS (Strater & Lipford, 2008). First hints for target memory problems in this context come from an exploratory interview study showing that Facebook users remembered the content of their self-disclosures well but struggled with remembering their corresponding privacy settings (Moll, Pieschl, & Bromme, 2014a). Thus, we predicted that SNS users should demonstrate substantial *target memory* problems (Hypothesis 1/ H1; for details see 5).

3. Metacognitive monitoring

Social Media users might be able to adequately compensate for their target memory problems if they became accurately aware of them. Thus, they need to metacognitively monitor (Nelson & Narens, 1994) their own target memory. Up to now, metacognitive monitoring has mostly been studied in contexts such as memory, learning, or decision-making (Dunning, 2012; Kessel et al., 2014). In a typical paradigm metacognitive judgments are collected as indicators of metacognitive monitoring, for example *retrospective confidence judgments* about the correctness of a retrieved answer (Nelson & Narens, 1994). For example, when answering forced-choice questions, a given answer is assumed to represent participants' best candidate answer among the given choices and a corresponding confidence judgment is assumed to represent the perceived probability of this answer being correct (Koriat & Goldsmith, 1996). To diagnose accuracy, such metacognitive judgments can be related to the corresponding performance (Pieschl, 2009): For example, *relative accuracy* is usually measured by correlations between metacognitive judgments and performance and refers to people's ability to discriminate between their own correct and incorrect answers.

Previous research revealed that people's metacognitive judgments do not reflect privileged access to their own cognitions but seem to be inherently inferential (Koriat, 2008, 2012). *Top-down* preconceived beliefs about one's own competencies and tasks can influence confidence judgments directly and also constrain the perception of bottom-up experiences during task solution (Dunning, 2012). For example, people are generally biased towards overconfidence, especially people with limited competencies (Kruger & Dunning, 1999), and especially regarding difficult tasks (Kessel et al., 2014). From a bottom-up perceptive, metacognitive judgments can be based on *information* or *experiences* (Koriat & Levy-Sadot, 1999). Information-based metacognitive judgments are educated as cues such as the number and quality of reasons for one's decision are deliberately processed; experience-based judgments resemble feelings as they directly utilize mnemonic cues derived from online task performance (Koriat, 2008). For example, the familiarity of stimuli, the accessibility of answer options, the speed of retrieved responses, and the consistency of answers positively impact confidence judgments (Dunning, 2012; Koriat, 2008, 2012). For example, Metcalfe, Schwartz, and Joaquim (1993) showed that confidence in cued recall performance was based on the familiarity of cues rather than on memory performance. Thus, metacognitive judgments are accurate as long as the utilized cues are valid and reliable.

However, in SNS communication (just as in FtF communication) many cues could be misleading. For example, most young people use SNS frequently and fairly competently which might result in preconceived beliefs about their own pronounced digital media literacy. First hints for metacognitive monitoring problems in this context come from an exploratory interview study showing that Facebook users' confidence judgments were not significantly related to their disclosure-related memory (Moll et al., 2014a). Thus, we predicted that SNS users should demonstrate substantial *metacognitive monitoring* problems when judging their target