Keller, T., & Savarimuthu, B. T. R. (2015). The Folly Guard: A personal and social communication protection system. In P. Kommers & P. Isaias (Eds.), Proceedings of the 13th e-Society International Conference. (pp. 233-238). IADIS Press.

THE FOLLY GUARD – A PERSONAL AND SOCIAL COMMUNICATION PROTECTION SYSTEM

Thomas Keller
Zurich University of Applied Sciences
Switzerland
Bastin Tony Roy Savarimuthu
University of Otago
New Zealand

ABSTRACT

The society has irrevocably changed. Just walking around a city makes this apparent. Single or in a group, people are accompanied by at least one electronic device which is able to communicate. The communication on the fly is intense and the communication industry is a multi-billion dollar marketplace ranging from tele communication infrastructure, communication products and services (e.g. social networks, mails and blogs). Information is produced at a rapid rate resulting in information overload. Because of the abundance of information, it becomes paramount that individuals find the right information at the right time. It's the relevance and salience of the information that makes it precious. Location based services or more general context based services are a means to find the relevant and salient information easier, with a higher degree of trust. At the same time the chances to get infected by malware is omnipresent. Most information systems are complemented by security modules that try to isolate different types of malware and to identify spam. The better the protection gets the less cautious the users become. They don't seem to realize that they expose themselves to these risks by using all the different communication means available. Even though the range of technologies used for communication has rapidly evolved, what has little changed is the psychology of the users. The evolution of human kind to use modern ICT technology took not just dozens of years but thousands. Hence, our brain may adapt on a rational and cognitive level to a certain degree but on a subconscious level we still react how we would have reacted hundreds of years ago. Psychologists' experiments show good examples like the priming effect or framing effect where individuals react instantly without deeper thought and regret the folly later. This instinctual subconscious behavior gains more relevance in a highly digital society where information is received and sent easily and quickly. The Folly Guard system presented here aims at identifying subconscious behavior as evident in the communication medium (emails, chat logs etc.) and warning users about the adverse effects which a user might not be unaware of. The Folly Guard thus operates at a psychological level to extract subconsciously created pitfalls in communication and expose those to the users. Thus, the system is meant to complement other IT security and protection systems, by operating in a domain that has not been previously explored. By potentially guarding against the follies of individuals, the system creates a social protection (i.e. minimizes the impact of follies at a societal level).

KEYWORDS

Security, Psychology, Information

1. INTRODUCTION

Imagine a project manager in his daily routine work. He will receive many mails, at the same time hurrying from one meeting to another (Carr, 2010). He receives a lot of information and disseminates quite as much. The time pressure is omnipresent and the goal must be to answer most mails within a tight time frame in order to keep to the various project time lines. The more stressed and emotional the work gets there are more chances of communication mistakes to happen. Looking back at the emails, sometimes one may shake their head in disbelief. Let's now have a look at the psychological side. Stanovic (2007) came up with a model of a human being consisting of two systems. System 1 represents our unconsciousness and intuitive abilities and system 2 stands for our rational and cognitive side. Generally, system 1 runs all the time. There is nothing like an on/off switch. System 2 on the contrary is potentially lazy and is triggered by system 1 when it comes to difficult situations where it needs system 2.

The efficiency of both system 1 and 2 is high and they normally produce an optimized performance which should make the *homo economicus* happy. The cooperation between the two systems works well most of the time. However, system 1 has its limitations. To overcome these limitations, system 2 needs to come into action. However, since it is lazy it must be actively triggered by system 1 and this doesn't happen all the time. Especially at special situations with a high emotional pressure, with a depleted system 2 or with a high time pressure situation, system 2 may not be able to reflect on the doings of system 1.

System 2 complements system 1 with its ability regarding logic and statistic. Under normal conditions it will wake up whenever it detects that system 1 requires support. System 2 has a natural speed. If it is forced to work above that its abilities are impaired. For instance, logical considerations are not coherent anymore. An example for system 2's laziness is the following puzzle (Kahneman, 2012):

A bat and ball cost $1.10.
The bat costs one dollar more than the ball.
How much does the ball cost?

The correct answer is 5¢. However, intuitively you may have come up with the solution 10¢. If so, your system 2 was just too lazy to come into action and to reflect on the intuitive reasoning of system 1. This may have happened at the end of a long and hectic day. The above puzzle might have been the last e-mail you just wanted to answer. Your depleted system 2 was not able anymore to react on system 1's intuitive answer. Once the mail has been sent, the wrong answer is already disseminated amongst your project members.

According (Kahneman, 2012) "…many thousands of university students have answered the bat-and-ball puzzle, and the results are shocking. More than 50% of students at Harvard, MIT, and Princeton gave the intuitive—incorrect—answer. At less selective universities, the rate of demonstrable failure to check was in excess of 80%. The bat-and-ball problem is our first encounter with an observation that will be a recurrent theme of this book: many people are overconfident, prone to place too much faith in their intuitions. They apparently find cognitive effort at least mildly unpleasant and avoid it as much as possible."

Let us perform another check on your system 2's alertness. Given the first two premises below, will the conclusion in the third sentence hold?

All roses are flowers.
Some flowers fade quickly.
Therefore some roses fade quickly.

Most certainly your system 2 was alert after the bat-and-ball example and found out where the flaw is. System 2 would have argued that roses may not belong to the set of flowers that fade quickly since only some flowers fade. System 2 in this case has overridden the intuitive answer from system 1 that states that the conclusion is correct. But again, imagine a special situation where you are under time pressure or depleted. In that case, it could be possible that your system 2 hasn't reacted on system 1's answer. According to Kahneman (2012) a large majority of college students endorse this syllogism as valid. And to be honest, after two premises which are obviously true and probably resemble our own experience it is hard declining the following conclusion. It definitely needs the logical reasoning of system 2.

Some people are more like system 1, e.g., impulsive and intuitive; others are more like system 2, e.g., capable of reasoning and are cautious. But the behavior is also dependent on the context and personal situation. These examples are situations where information systems can support the user by identifying a potential critical situation and provide a warning together with some useful hints.

2. PROPOSED USE CASES

The following sub-sections outline four possible uses cases for the application of a social protection system that monitors communication of an individual user (both received and composed messages) and warns of potential pitfalls from a psychological stand point (i.e. and possible pitfalls of system 1). More examples can be found in Kahneman's book (Kahneman, 2012).

2.1. Priming

Making a suggestion in a conversation is an example of priming. For instance if you are asked to guess the average cost of a car, e.g., in Germany, along with a suggestion of a high average cost you end up remembering luxury brands (Strack, 1997). Another example is the following two questions (Kahneman, 2012): "Is the height of the tallest redwood more or less than 1,200 feet? What is your best guess about the height of the tallest redwood?"

The suggestion of 1200 feet is ridiculously high. Even if this obviously wrong suggestion is identified as implausible, it has an influence on your answer for the second question. Experiments have shown that different groups of people which have been offered different suggestions for the height for the first question came up with considerably different guesses for the second question.

An excerpt from an email from someone in the project management office may read: "…usually the test phase of a project of such limited complexity does not take more than 8 days. What is your best guess of the test duration?"

It is apparent that the sender of this mail is determined to set an anchor for the duration of the test phase. Usually there is always little time left for testing purposes. So, it is most likely that the proposed 8 days are insufficient for a decent test procedure. Galinsky (2001) has proposed subtle ways to resist the anchoring effect. It mainly requires system 2 to come up with arguments against the anchor. In the context of the email example, the receiver of the mail would need to come up with examples of projects with considerably longer test durations to counter the proposed duration.

This is the place where information systems can help. A social protection system that operates on top of an email client can point out that the numbers mentioned in the communication thread must be verified against some statistical data. To do so, the protection system needs to have access to additional data stores to retrieve adequate information.

2.2. A Bias to Believe and Confirm

According to Gilbert (2007), the understanding of a statement must begin with an attempt to believe it. You must first know what the idea would mean if it were true. This approach is automatically performed by system 1 and it comes up with a construction of a well-fitting interpretation of the story or of a possible cause and effect. If system 2 doesn't come into action now the statement is taken as true. It needs system 2 challenging the statement and constructing a cause-effect that opposes the statement. Gilbert (2007) sees unbelieving as an operation of System 2 which may be supported by additional information offered by ICT systems. The additional information might be needed to construct different cause and effect and is a means to get system 2 to reason.

2.3. What You See Is All There Is (WYSIATI)

An essential design feature of the associative machine (system 1) is that it represents only activated ideas. Information that is not retrieved (even unconsciously) from memory might as well not exist. System 1 excels at constructing the best possible story that incorporates ideas currently activated, but it does not (cannot) allow for information it does not have. The measure of success for system 1 is the coherence of the story it manages to create. The amount and quality of the data on which the story is based are largely irrelevant. When information is scarce, which is a common occurrence, system 1 operates as a machine for jumping to conclusions.

As the WYSIATI rule implies, neither the quantity nor the quality of the evidence counts for much in subjective confidence. The confidence that individuals have in their beliefs depends mostly on the quality of the story they can tell about what they see, even if they see little. We often fail to allow for the possibility that evidence that should be critical to our judgment is missing—what we see is all there is. Furthermore, our associative system tends to settle on a coherent pattern of activation and suppresses doubt and ambiguity.

Here again information systems can be of help. If a social protection system has the ability to identify situations featuring limited information it could come up with some additional context information that may have an impact on the coherence of the story produced by system 1, thus thwarting premature conclusions.

2.4. Framing

A framing effect refers to situations where the same information can be presented in different ways evoking different emotions. The statement that "the odds of survival one month after surgery are 90%" is more reassuring than the equivalent statement that "mortality within one month of surgery is 10%." (Kahneman, 2012). Similarly, a project manager pointing out that "95% of the goals have been achieved" is generally taken as a positive message. The equivalent formulation "5% of the goals have not been reached" contains a negative message and will presumably invoke rather negative feelings. The equivalence of the alternative formulations is transparent, but an individual normally sees only one formulation, and what he sees is all there is.

This is again a perfect scenario for the use of social protection system. To provide a balanced view to the user about an issue, as a first step, the protection system might not even need to come up with the equivalent formulations. Just pointing out that there is an equivalent formulation will help to realize the framing effect.

3. THE SOCIAL PROTECTION SYSTEM

In this section, we outline a possible design of the Folly Guard system and also point to the architectural options. The Folly Guard is a software component that has the ability to identify situations where system 2 should have been activated but for whatever reason has not come into action. The basic functionality of Folly Guard is the analysis of a thread of communication and the provision of warnings and useful hints with the aim to activate system 2 of the concerned user.

3.1. Folly Guard Design

The design of the Folly Guard is depicted in Figure 1. The core element of the system is a natural language processor (Jurafsky and Martin, 2008; Baldrige, 2005; Bird, 2006) that takes as input one or several communication threads and analyzes it accordingly (Manning et al, 2014). The results are then sent to the plugins which try to identify potential psychological pitfalls. The produced outputs are warnings and hints about possible pitfalls which should be scrutinized by the user. The overall functionality is defined by the amount and type of plugins available for the processor. It is assumed that for the each psychological pitfall a separate plugin takes responsibility (e.g. plugin 1 for identifying priming and plugin 2 for identifying framing). This approach makes the system flexible and adaptable for future enhancements.

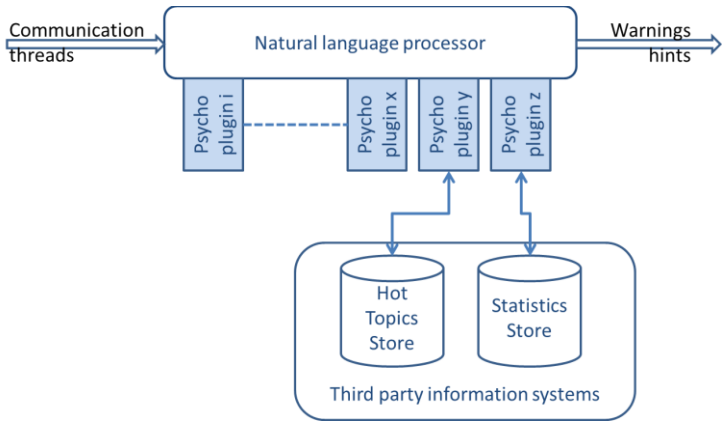

Figure 1: Design proposal for the Folly Guard including possible 3rd party datastores

For instance for the affect heuristic (Slovic et al., 2007) it may be necessary to have more information than is available within the communication thread. Therefore, there must be data stores accessible to query for more information based on the current context of the communication thread under analysis. Figure 1 gives two

examples for possible external data stores. We note that the Folly Guard system's design allows for personalization. That means the Folly Guard could maintain a history of detected pitfalls, associate it to a specific user and learn from them by inferring norms.

3.2. Architectural options

Architecturally, the Folly Guard system can be developed as a client-server system as shown in Figure 2, where the Folly Guard module (referred to as Psycho Guard in Figure 2) can be placed on the client side or the server side. In the client side, the module for Folly Guard can be delivered as a plugin (e.g. plugin for the email client to analyze emails) or a client side proxy or a wrapper to a legacy client. At the server side, it can be a pure server module or a proxy.

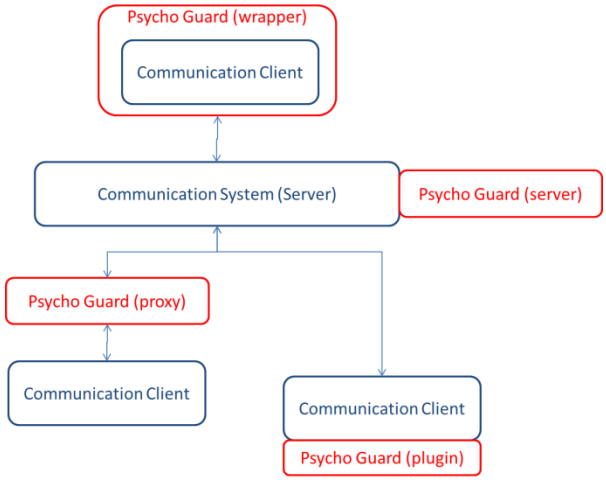

Figure 2: Architectural options for the Folly Guard system

4. DISCUSSION

To date, prior research hasn't addressed the need for guarding individuals against follies when they communicate. Marketing researchers have studied how traditional media primes people (Roskos-Ewoldsen et al., 2000). Recently, the nature of priming in online social media has also been studied (Doyle, 2015). Researchers have investigated how priming can be inferred from presented sentences (McKoon and Ratcliff, 1984). However, there hasn't been an attempt to design a system that can help in the identification of several aspects of communication (e.g. priming and framing) that contain potential psychological pitfalls. Upon identification, these can then be presented to the user so as to minimize harmful impact.

Based on the premises of a human being consisting of the two systems 1 and 2, an omnipresent work pressure and an increasingly faster digital communication there is a demand for some support to avoid some common psychological pitfalls. A solution to this is the Folly Guard system. The system identifies potential pitfalls from communication channels and presents those to the users, thus acting as a proxy for human's system 2.

The technological cycles are much faster than our human brain can adapt to. Since the Folly Guard is not meant to alter any information there is no danger of potential changes to communication content without prompting the user to acknowledge it. Rather, Folly Guard comes up with some hints the user may take serious or not. But, in the end there is hope to improve communication and to lessen communication mistakes.

To the best of our knowledge there haven't been surveys of requirements of potential users for such a system. An interesting cohort to consider for eliciting requirements would be members participating in a professional project (e.g. a distributed software development project that requires extensive communication). Also lower management and HR would be good user groups to consult to obtain requirements as they are likely to have experienced common follies associated with communication. Based on the requirements collected, a first draft of the design will be worked out. We currently envisage the development of prototype Folly Guard for an email system since it might be the easiest to implement. A good choice for the first psychological construct to extract from free text is the priming effect. Subsequent to the construction of the plugin, efficiency of identification (e.g. using precision and recall metrics) must be measured and also a field test must be carried out to assess the usefulness of the system. Other potential applications where the Folly Guard module can be integrated are a) app clients for Twitter b) SMS clients c) and clients of social networking applications such as Facebook. This will help avoid common psychological pitfalls in communication, both for individuals and organizations. We believe such a system when adopted *en masse*, will help reduce the amount of psychological pitfalls encountered by the whole society.

REFERENCES

Baldridge, J., 2005, *The Apache OpenNLP project*,

Bargh, John A.; Chen, Mark; Burrows, Lara, 1996, *Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action*, Journal of Personality and Social Psychology, Vol 71(2), Aug 1996, 230-244

Bird, S., 2006, *NLTK: the natural language toolkit*, In Proceedings of the COLING/ACL on Interactive presentation

Galinsky, A. D.; Mussweiler, T., 2001, *First offers as anchors: The role of perspective-taking and negotiator focus*, Journal of Personality and Social Psychology, Vol 81(4), Oct 2001, 657-669.

Carr, N., 2010. The Shallows. What the Internet is doing to our brains.

Doyle, E., 2015, "Context, Context, Context: Priming Theory and Attitudes Towards Corporations in Social Media", Discussion Paper, Accessed from http://www.academia.edu/6885596/Context_Context_Context_Priming_Theory_and_Attitudes_Towards_Corporations_in_Social_Media, 14 January 2015.

Gilbert, D., 2007, How Mental Systems Believe, *American Psychologist*, 46(2), 107-119

Gilbert, D., 2007, *Stumbling to Happiness*, Vintage

Jurafsky, D., Martin, J. H., 2008, *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*, 2nd edition. Prentice-Hall.

Kahneman, D., 2012, *Thinking, Fast and Slow*, Penguin Press, UK

Manning, Ch. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, St. J., and McClosky, D., 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.

McKoon, G., & Ratcliff, R., 1984. Priming and on-line text comprehension. In Kieras & Just, (Eds.), New methods in reading comprehension research (pp. 119-128). Hillsdale, N.J.: Erlbaum.

Pólya, G., 1945, *How to Solve It*, Princeton University Press

Roskos-Ewoldsen, D. R., Roskos-Ewoldsen, B., and Carpentier, F.R.D., 2002, "Media priming: A synthesis" Media effects: Advances in theory and research 2: 97-120.

Selke, A., 2006, Informationsmanagement in schnell wachsenden Unternehmen (Vol. 4). kassel university press GmbH.

Stanovich, K., E., 2011, *Rationality and the Reflective Mind*, Oxford University Press

Strack, F.; Mussweiler, T., 1997, *Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility*, Journal of Personality and Social Psychology, Vol 73(3), Sep 1997, 437-446.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G., 2007, *The affect heuristic*, European Journal of Operational Research, 177(3), 1333-1352.

Wilson, T., D., 2014, *Strangers to Ourselves*, Belknap Press sessions (pp. 69-72). Association for Computational Linguistics.