The Comic Strip Game:
observing the impact of implicit feedback in the content creation process

Fiammetta Ghedini and **Benjamin Frantz** and **François Pachet** and **Pierre Roy**
SONY CSL Paris
fiammettaghedini@gmail.com

Abstract

The Comic Strip Game is a system allowing users to create dialogues for speechless cartoon strips during shared, online content creation sessions. This paper describes the results of a protocol providing each participant with implicit feedback and inspiration from other participants. We observed the behaviour of subjects and investigate the impact of other participants' behaviour on their creative process.

1 Introduction

We posit that a creation task involves a succession of production and self-evaluation steps, which, usually, are not explicit and thus cannot be observed [Lubart 2013]. We have designed an experimental protocol in which production and evaluation steps are explicit, as well as the influence that other people work can have on one's own. Specifically, the objective of the Comic Strip Game is to investigate (1) positive or negative bias in implicit self-evaluation of creativity, (2) the impact of implicit feedback from other participants, (3) the attachment to one's own creation and (4) the existence of consistent content creation strategies and their distribution on the subjects' population.

In order to do so, we designed an online system proposing ten different cartoon strips (see all the proposed strips in Annex I, at the end of this document). A strip is a series of images separated from each other, each image representing a character or a scene happening in a visual unity called "panel". In each panel of the strips, one or two characters are talking, by means of the traditional symbol of the "balloons", which were left blank, for the participants to write up. The content creation task is therefore to invent a story for the strip and to express it via the text in the balloons. In the Comic Strip Game, this content creation task is neither solitary nor collaborative: it is rather concerned with mechanisms of implicit feedback from other participants, as detailed in the protocol in the section below.

1.1 Motivations and background

Evaluation, both external and "internal" (self-evaluation) is a central issue of any creative process. The importance of self-evaluation and its role on the creative process itself, nevertheless, have not been extensively studied, probably more because of the difficulty of defining and isolating it with more subtle means that administrating questionnaires. This is why we thought that designing a system able to provide implicit feedback and imposing several implicit self-evaluation steps may be a valid methodology in order to investigate the creative process.

In general, previous literature has discussed whether the potential for external evaluation can affect the creative process; the first study by [Amabile, 1979], confirmed by [Bartis et al, 1998] highlighted a decrement in creativity due to external evaluation. [Szymanski & Harkins, 1992] partly confirmed the harmful effect of external evaluation on creativity (but not on the performance itself) during the process of generating as many uses as possible for an object. [Silvia and Philips 2004] suggests that also self-evaluation reduces creativity (for tasks involving generating remote associates and finding unusual uses for objects).

From another point of view, as external assessment of creativity can be taken into account when judges reach an agreement, the validity itself of self-evaluation in creativity has been questioned. For instance [Kaufman et al 2010] compared self-reports of creativity in four artistic domains to experts' judgments: external and self-evaluations did not correlate. In a similar study applied specifically to music, [Priest 2006] compared students' self-assessments of musical compositions and experts' assessments. In this study too, there was no significant correlation between the judges' evaluations and the students' reports.

On the other hand, we do not refer to the theoretical framework of collaborative creativity, which presents specific characteristics, such as "idea talk", variance in contributions, roles artificial or spontaneous attribution [Freeman 2014] which are made impossible by the constraints imposed by our system, for which the feedback has to remain implicit.

Indeed, the design of our study is meant to investigate the choices and evaluations present in the creative process bypassing an explicit self-evaluation step, by presenting the subjects with implicit feedback from other participants. Implicit feedback that every subject will receive from co-participants should also deflect the social loafing phenomenon, which is the tendency for individuals to lower

their productivity when in a group (see [Simms 2014] for an extensive review and [Williams et al. 1981] for how potential evaluation decreases social loafing.

1.2 Methodology

The experiment takes place online, where four subjects are randomly assigned to create dialogues for the same strip. As explained above, each strip consists in a set of images made of three to four panels (See Annex I), in which balloons are left blank. Each subject can complete from one to eight different stories.

Before the content creation session begins, subjects are required to declare their age, gender, mother tongue and all the languages they are fluent in. We then form groups of four subjects who will participate for a single session. The groups are formed randomly but each member has in common the knowledge of at least one language (the one in which the dialogues are going to be written) and the fact that it is the first and only time they created content for that specific strip. Subjects assigned to the same strip are anonymous and cannot directly interact with each other. Before the session begins, subjects are instructed on the task they will perform by means of a tutorial, with a particular emphasis on the fact that the objective is not only to write but also to choose the best story (even if the subject is not the author of it). To reinforce this concept and motivate the subjects, we stated that one of the participants would be chosen randomly to win a SONY tablet, among the participants who would choose (thus not necessarily write) one of the best stories.

During each stage of the content session creation, the subjects are producing dialogues (text) for one panel. When all the subjects have written and submitted their first dialogue, their texts are proposed to all four participants, including themselves. Subjects are instructed to choose other participants' proposals if they evaluate they have a greater potential for further developing the strip, therefore subjects, at this stage, can decide to pursue the story by selecting their own text proposal or to switch to the proposal of someone else. The same protocol applies for the second, third and fourth step.

We refer to the behaviour of choosing another subject's' proposal as "switch", therefore to subjects performing it as "switchers". We refer to the subjects who choose to continue their own production as "pursuers".

The final step consists in asking subjects to choose the story they evaluate as the best one. As explained above, before the beginning of the experiment subjects were instructed to choose the best one, regardless of how much they contributed to it.

For the sake of clarity, since the protocol is quite complex, we describe hereafter the typical development of one session (dialogues and behaviours come from our actual dataset, but names are fictional).

An example session from the Comic Strip Game

Vincent, Paul, Francesca and Benoit participate to their first session. Paul is Dutch, Francesca is Italian, Vincent and Benoit are French, but they are all fluent in English. They do not personally know each other.

The four subjects state their gender, age and spoken languages before they begin to play. When all these data is entered, they visualise the strip which the system has randomly chosen for them (the only condition is that nobody among them has already created a dialogue for this strip). The strip depicts, in this case, a ginger cat in different positions.
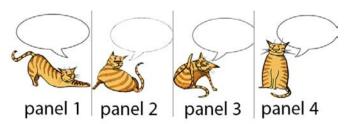

Figure 1: One of the proposed strips

They all propose a text for the first panel (See Fig.1), which is automatically inserted in the balloon.

Vincent proposes: Nothing like a good night
Paul proposes: Oh, c'est parfois difficile de se lever, le matin!
Francesca proposes: The Smooth Slope…
Benoit proposes: Usual stretching and we can start

They submit their proposal by clicking on a "send" button, and when everybody is done, they are able to read the proposal written by others. Now they can either choose to pursue their own story (that is to say, to write a text for Panel 2 by continuing with their own text) or to switch to the proposal of someone else. This is what happens:

Vincent chooses to "abandon" his text and switches to the text of Benoit: this means that during the next step, he will have to write up the second panel, as a continuation of "Usual stretching and we can start".
Paul, who probably got confused with the experiment's language as he has been writing in French, makes the same choice as Vincent: he switches in favour of Benoit's text.
Francesca keeps her text.
Benoit decides to switch and choses Vincent's text. Now they propose a text for the second panel (the first panel's text in italics):

Vincent proposes: *(1) Usual stretching and we can start* (2) It's important to be well prepared
Paul proposes: *(1) Usual stretching and we can start)* (2) In this position, I look quite fat!
Francesca proposes: *(1) The Smooth Slope* (2)…the Easter Egg…

Benoit proposes: *(1) Nothing like a good night* (2) Sorry for the back.

Now the subjects choose again the chain of two panels to be continued. It is worth mentioning that during this step (before proposing a text for the third panel) the subjects are exposed to the first implicit feedback from other participants, because they realise whether their first proposed text has been chosen by other subjects (and therefore its creative potential has been implicitly evaluated as good) or not. For instance, now Benoit, who abandoned his own proposal, realises that his text has been chosen by two other participants, Vincent and Paul. Vincent, once again, switches in favour of Benoit. He does not know who is submitting the text, as participants are invisible to each other and do not even have a nickname, so there is no possible bias in choosing a specific author: what is chosen is always and only the text. Paul switches in favour of the story of Francesca. He likes her idea (which he had probably not understood at the beginning): the cat is doing yoga. Francesca keeps her text. Benoit, this time, decides to keep his text.

Now the subjects propose a text for the third panel (in italics the first and second panel):

Vincent proposes: (1) Nothing like a good night) (2) Sorry for the back (3) Wow. I didn't know I could do that with my leg!

Paul proposes: (1) The Smooth Slope… (2)…the Easter Egg… (3) The One Beer for me…

Francesca proposes: (1) The Smooth Slope… (2)…the Easter Egg… (3) …the antelope…

Benoit proposes: (1) Nothing like a good night (2) Sorry for the back (3) First task: bathing!

The subjects once again chooses their preferred storyline:

Vincent chooses Paul's text.

Paul keeps his text.

Francesca choses Paul's text.

Benoit keeps his text.

Now they propose a text for the fourth panel, the "punchline" of the story.

Vincent proposes: *(1) The Smooth Slope… (2)…the Easter Egg… (3) The One Beer for me…* (4)… Yoga for cats

Benoit proposes: *(1) Nothing like a good night (2) Sorry for the back (3) First task: bathing!* (4)… And also for today, I am done working!

In the final step, the subjects have to vote for the best story: the consensus is complete, as everybody votes for Paul's ending.

Figure 2 illustrates the whole process visually:

Figure 2: the process of creation and choice, on 5 steps. Each colour corresponds to a player.

2   Results

We recorded, at each step, whether the subjects have pursued their own story or switched to the story of someone else. In addition to this "switching behavior", we also recorded the number of votes each participants received at each step. This measurement was used as a quality level of each text, and the mean number of vote received as a performance level for each participant.

2.1   Description of the population

Among the 953 individuals who registered to the experiment's website, 756 subjects did not complete a single session and were excluded from analysis, leading to a sample of 197 subjects who completed at least one session. The high difference between registered users and actual subjects may be due to the fact that seldom four potential subjects sharing the same language and the same strips to be completed would be online at the same time.

Figure 3 illustrates the number of participants according to the number of experiences completed, where it can be observed that the majority of subjects concluded only one session. This may be due to the length of the waiting time between sessions to find other participants or to the perceived difficulty of the task.

Among this population, the mean age is 33.7 years, with a standard deviation of 11.9 years. The youngest subject is 11 years old and the oldest one is 73.

Gender is equally distributed among the subjects, with a 53% of males and 47% females.

The experiment was available in five different languages (English, French, Italian, Dutch and Spanish) and the distributions of the native languages and fluent languages are illustrated by Figure 4.
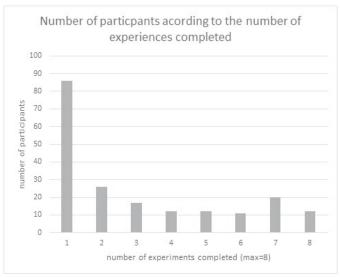

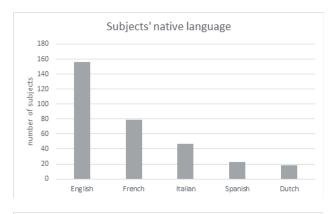*Figure 3: attendance to sessions*

Figure 4: Distribution of native and spoken languages among the population

2.2 Emerging profiles: switchers and pursuers

Table 5 illustrates the subjects' behaviour during each step implying a decision (switching in favour of another person's text or pursuing their text), thus identifying the emerging "profiles types". It occurs that "extreme" profiles are the most common ones: the most frequent behaviour corresponds to the profile type 1, or the "perfect pursuer",

Profile type	Switching behaviour	Number of subjects (mean profile)	Number of subjects (exact profile)
1	0000	44	20
2	1000	25	14
3	0100	10	2
4	0010	7	1
5	0001	5	1
6	1100	6	2
7	1010	5	3
8	1001	4	3
9	0110	4	3
10	0101	2	1
11	0011	1	0
12	1110	17	6
13	1101	9	3
14	1011	5	2
15	0111	4	3
16	1111	48	17
	Total	196	81

Figure 5: profile types and their distribution among the population. In the second column, 0=pursue and 1= switch

who never switches for another subjects' content. It is followed by the "perfect switcher" profile, who keeps abandoning her/his own production in favour of the production on someone else. In third position (see Figure 6) we find those subjects who switch only once, during the first step, and then pursue their production. Because subjects can participates to several experiments (up to eight) their profiles are not necessary systematics. For example, a subject could choose not to switch during the first experiments, but could choose to switch during the others.
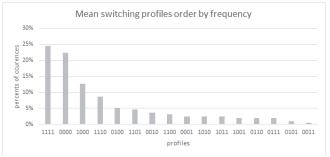

Figure 6: The most frequent profiles

We observe (Figure 7) that the three-step profile's distribution is very similar to the distribution taking into
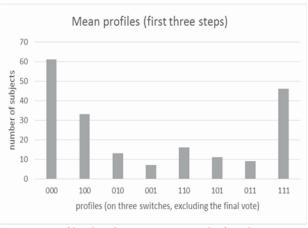

Figure 7: Profiles distribution concerning the first three steps

account the four steps: the most frequent profiles are the perfect switchers and the perfect pursuers, followed by the ones who switch only during the first step. Theses results suggest that the profile distribution is consistent, reliable and is not an artefact due to a windowing effect.

2.3 Switching stability

By observing the total switching rates, we can determine that the switch and pursue rates are coherent and constant throughout the different experiments. This means that the number of sessions to whom each subject participated does not impact the switching rates, as illustrated in Figure 8.
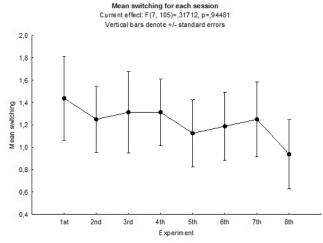
Mean switching for each session
Current effect: F(7, 105)=,31712, p=,94481
Vertical bars denote +/- standard errors

Figure 8: the switching mean remains stable throughout the eight experiments.

On the other hand, we can see that the number of switches significantly decreases inside the sessions: subjects switch less and less at each evaluation step ($F[3,522] = 11,084$, $p<.001$), as illustrated in Figure 9:

Mean switching according to the rank of proposition
Current effect: F(3, 522)=11,084, p=,00000
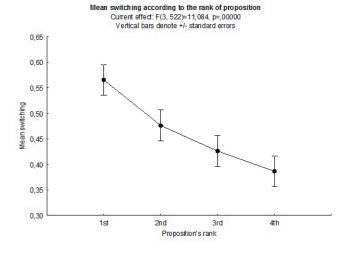Vertical bars denote +/- standard errors

Figure 9: subjects' tendency to switch significantly decreases at each step

2.4 Votes

As mentioned above, subjects can either switch or continue their own text, knowing that the very text they produced can be chosen by other participants, whether or not the author is a switcher. We refer to the event of a text being chosen by someone else as a "vote".

Figure 10 illustrates the mean of votes received for each step by the "pursuers" (the 50% subjects who pursue the most in light grey) and by the "switchers" (the 50% subjects who switch the most in dark grey).
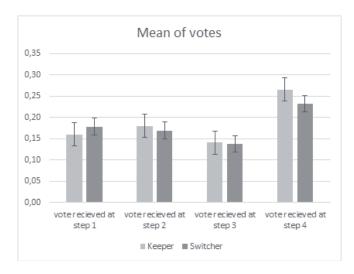

Figure 10: votes received by pursuers and votes received by switchers

Votes received by pursuers and votes received by switchers are not significantly different ($F[8,196]=1,390$, $p>.10$). This results stays true even on more contrasted groups such as the "perfect switchers" and the "perfect pursuers".

The number of votes received by a text which has been "abandoned" by its author (mean =0.87) is not significantly different from the number of votes received by a text which has been kept by its author (mean = 0,76) ($t[164] = 0.790$; $p = .43$). This result stays stable for each step.

2.5 Returners

We can identify another common behaviour, the one of the "returners". The returners are those switchers who abandoned their first and/or second panel's position productions and recover them if another subject "adopts" them. Among the 487 stories where the first panel's text was abandoned by its author, only 130 (27%) were abandoned also by everybody else . On the 357 stories remaining, (where, therefore, a "return" was possible), we observed 109 (31%) returns, 165 (46%) continuations where the original author did not return and decided to pursue his/her production and 83 (23%) switches where the original author did not return and decided to switch again for the story of another subject.

Among all texts the repartition of returns is as follow: 14% have been abandoned at step 1, voted by one or more

subjects and then recovered at stage 3 by their original authors (returns); 5% have been abandoned at step 1, voted by one or more subjects and then recovered at stage 4 by their original authors; 15% have been abandoned at step 2, voted by one or more subjects and then recovered at stage 4 by their original authors. This results show that the returns are more likely to occur when the initial switch (abandon) and the return are separated by only one panel.

2.6 Gender and language effect

Results show no effects between male and female participants, whether it is for the switching rates (t[191] = 0.40 ; p = .31) or the number of votes received (t[191] = 1.17 ; p = .22).

Because the task was available in five different languages, we also tested the effect of language and found no significant differences between the number of spoken languages and the switching rates (F[4,191] = 1.03 ; p = .39) or the number of vote received (F[4,191] = 1.37 ; p = .24). No differences were observed also between the participants' mother tongue and the switching rates (F[4,187] = 1.38 ; p = .24) or the mean number of vote received (F[4,187] = 0.14 ; p = .96).

3 Discussion

3.1 Self-evaluation bias or engagement?

In accordance to previous studies [Kaufman and Evans 2010] [Priest 2006], our findings challenge the validity of self-assessments in creativity. This result stems from the decrease of the switching rate for the later steps, which is significant even if the subjects were instructed and motivated to choose the best story independently from their contribution to it: this means that, at the end, a majority of subjects judged their own story as the best one. The motivations for this observed behaviour could be a selfenhancing bias and/or a progressive engagement in one own' work.

The self-enhancing bias, or self-serving bias, is the tendency to perceive oneself more positively than a normative criterion would predict [Krueger 1998]. This could be the explication of the significant tendency to chose one's own creation as the best one in the final steps. Regarding the outcomes of the first steps, apparently in contradiction with the self-enhancing bias theory, they are consistent with the exploratory behavior usually observed in creativity tasks [Finke, Ward & Smith, 1992]. Aself-serving bias specific to the creative process is a result that has never been highlighted before, to our knowledge.

Another motivation for the significant decrease of the switching rate in the later steps may also be the increasing effort that subjects have applied in the creative process. The design of the experiment itself, proposing a sequential creative activity, makes possible to highlight this occurrence. Subjects are indeed more open to switch at the beginning of the process, but as they put effort and invest their time in the task, they become more attached to their production, as a commitment effect [Beauvois, Joule & Brunetti, 1993]. This result also implies that it is easier to change the direction of a creative work at its early stages rather than towards the end. To our knowledge, this result has never been captured by a scientific experiment; nevertheless, [MacKinnon 1978] highlighted that experienced architects are more likely to abandon their ideas than beginners. This could mean that, independently from its motivation, being aware of the bias "against change" is a skill of the creative professional.

3.2 Profiles and quality of the outputs

We could not observe any significant impact of the profile type on the quality of the productions, evaluated by the number of votes received (F[7,188] = 1.03, NS). We thus tried to analyse patterns in the ten strips that received the most consensus. Four of them received a unanimous consensus from all four participants, and six from three out of four. We determined that three of them were composed by only one author who never switched (and it should be noted that two of these stories were made by the same author, indicating a very creative participant), six were composed by two different authors and one by three different authors. This result [FG6] suggests that consensus can be more easily attained when stories are created through collaboration (using or giving ideas from/to others). Interestingly, we can see that on the seven experiments where there was a consensus of three participants, the nonconsensual response by the fourth participant was on six times out of seven for his or her own production. This reinforce the result that the tendency of selecting one's own story during the final step is so strong that participants prefer to do so even when there is a worthwhile story. We can also observe that the profile distribution of the subjects who has drawn the most votes from their coparticipants is the same as the whole sample. This confirms that the switching or pursuing profile is not linked to the quality of the productions.

Another interesting result concerning the profiles is that the most frequent profiles are "extremes", that is participants who never switch or always switch. One explanation would be in term of personality, mainly the openness dimension with its tendency to explore other ideas, to try something new. Another one would be in term of self-esteem, where subjects with a low self-esteem would consistently judge, and hence chose, other stories better than their own production, whereas subjects with high self-esteem will do the opposite. These suggestions are purely speculative, and it would be interesting to replicate this study with a personality questionnaire and a self-esteem evaluation to assess them.

3.3 Language effect

Interestingly, the results concerning language seems to contradict the classical advantage of multilinguals on creativity tasks, where they usually outperform the monolinguals [e.g., Karapetsas and Andreou, 1999;

Kharkhurin, 2008]. However, the absence of effect could depend from the self-evaluation, since productions were assessed by the groups themselves through the number of votes each text received.

3.4 Effect of implicit feedback and "returning" behavior

In the context of the Comic Strip Game, the subjects have to select which story they will continue. This implicitly indicates to the subjects that they will receive feedback on their production, not only at the end of the strip as mentioned in the instructions, but also at the end of each panel, in a within-group evaluation which should deflect, as explained above, most of the social loafing effect [Szymanski & Harkins, 1992]. Indeed, steps 3 and 4 provide the participants with implicit feedback on their previous productions because they can see if their text has been chosen or not by other subjects. This implicit feedback can be particularly interesting when a subject has abandoned his or her first production(s) and then realises at step 3 or 4 that someone else has selected his/her "abandoned" text. This can lead to a cognitive dissonance for the author of the first text because he or she judged it not good enough, or with less potential than other texts, but others saw instead a good idea or an interesting potential [Festinger, 1962].

We observed that authors who have abandoned their text at the first panel (i.e. they have switch at the first step) and could return to continue their story later, in a large majority they did not (69%). In other words, once an idea is discarded, it is for good and reconsidering it is less likely to happen. This result is consistent with the commitment effect described earlier which suppose a consistency from previous choices and a difficulty to change opinion or judgement, particularly when the choice was voluntary and not constraint. Moreover, we can see that this effect is more important when the delay between the first switch and the return is longer, because when the delay is one panel long (between first and third step or between second and fourth step), we have a constant return rate of 15%, while only 5% when it is two panel long (between first and fourth step).

These results suggest that the feedback provided by the votes of the other participants are not enough to compensate the "anti-change" bias.

3.5 Conclusion and future work

The data analysis of the Comic Strip Game has given us an insight on the creative process. Our results highlight that the potential impact of implicit feedback from other participants and objectivity in self-evaluation, even if encouraged, are lessened by a bias "against change". Such a bias probably stems from a combination of selfenhancing bias and of a commitment effect. We could also highlight consistent and stable strategies for content creation which, interestingly, are not related to gender, age and spoken languages.

Future work may focus on the "against-change bias", for example to test whether it resist to within group social pressure, and to explicit, external or internal feedback. It may be also interesting to design a protocol investigating the motivations of the bias, in order to distinguish between the self-serving illusion and the commitment effect.

References

[Amabile 1979] Amabile, T. M., Effects of external evaluation on artistic creativity, Journal of Personality and Social Psychology, Vol 37(2), Feb 1979, 221-233

[Bartis 1988] Bartis, S., Szymanski, K., & Harkins, S. G. (1988). Evaluation and performance: A twoedged knife. Pers Soc Psychol Bull June 1988 vol. 14 no. 2 242-251

[Beauvois 1993] Beauvois, J.-L., Joule, R.-V., & Brunetti, F. (1993). Cognitive rationalization and act rationalization in an escalation of commitment. Basic and Applied Social Psychology, 14(1), 1–17.

[Festinger 1962] Festinger, L. (1962). A theory of cognitive dissonance (Vol. 2). Stanford university press.

[Finke 1992], Finke, R. A., Ward, T. B., & Smith, S. M. (1992). Creative cognition: Theory, research, and applications.

[Freeman 2014] Freeman,O., Tangney, D., O'Rourke, B.K., (2014): Performing Collaborative Creativity: Learning from Diverse Experts Interacting in Ireland's Science Gallery, 30th European Group for Organizational Studies Colloquium, Rotterdam

[Karapetsas 1999] Karapetsas, A., & Andreou, G. (1999). Cognitive development of fluent and nonfluent bilingual speakers assessed with tachistoscopic techniques. Psychological Reports, 84, 697–700.

[Kaufman and Evans 2010] Kaufman, J. and Evans. M., The American Idol Effect: Are Students Good Judges of Their Creativity across Domains? Empirical Studies of the Arts January 2010) vol. 28 no. 1 3-17

[Kharkhurin 2008] Kharkhurin A. V. (2008). The effect of linguistic proficiency, age of second language acquisition, and length of exposure to a new cultural environment on bilinguals' divergent thinking. Bilingualism: Language and Cognition, 11, 225–243.

[Krueger 1998] Krueger, B, Enhancement bias in Descriptions of Self and Others, PSPB, Vol.24, No5, May 1998, pp. 505-516

[Lubart 2013] Lubart, T., Mouchiroud, C., Tordjman, S. & Zenasni, F. (2003). Psychologie de la créativité [Psychology of creativity), Paris, Armand Colin

[MacKinnon 1978] MacKinnon, D. W. (1978). In search of human effectiveness: Identifying and developing creativity. Buffalo, NY: Creative Education Foundation.

[Priest 2006] Priest, T., Self-evaluation, creativity, and musical achievement, Psychology of Music, January 2006, vol. 34 no. 1 47-61

[Silvia and Phillips 2004] Silvia P. and Philips A. 2004, Self-Awareness, Self-Evaluation, and Creativity, Pers Soc Psychol Bull August 2004 vol. 30 no. 8 1009-1017

[Simms 2014] Simms, A., (2014) Social Loafing: A Review of the Literature, Journal of Management Policy and Practice vol. 15(1)

[Szymanski & Harkins 1992] Szymanski, K., & Harkins, S. G. (1992). Self-evaluation and creativity. Personality and Social Psychology Bulletin, 18(3), 259-265.