

Prof. Agnieszka Wykowska
Istituto Italiano di Tecnologia
Social Cognition in Human-Robot Interaction
Via Enrico Melen 83, 16152 Genova, Italy
Tel: +39 010 8172242
Email: agnieszka.wykowska@iit.it
Website:
<https://www.iit.it/people/agnieszka-wykowska>

Prof. Dr. hab. Przemysław Bąbel
Przewodniczący
Rady Dyscypliny Psychologia
ul. Ingardena 6
30-060 Kraków

Genova, 12.05.2021

Evaluation of PhD thesis of Mr. Bartosz Majchrowicz
“Prediction, learning and signals integration in the sense of agency”
Jagiellonian University in Kraków

The PhD thesis of Mr. Majchrowicz focuses on examining various factors which influence implicit and explicit sense of agency. It addresses the question of how predictability, learning, and affective valence of the sensory outcomes of actions affect sense of agency. Importantly, one of the foci of the thesis is to examine the relationship between various measures of sense of agency – a very timely and crucial issue that needs to be addressed in the field abundant with various paradigms and measures.

The thesis comprises an Introduction section elaborating on the theoretical background, respective concepts and methodological considerations. This is followed by short summaries of the empirical work, which are then followed by three already published papers. In the first section, Mr. Majchrowicz introduces various aspects related to sense of agency. He structures the introduction according to three perspectives: phenomenological, methodological and mechanistic.

Phenomenological perspective deals with the experience of agency. Here, Mr. Majchrowicz distinguishes between implicit feeling of agency and explicit judgment of agency. The first one is related to the minimal self, namely a more pre-reflective phenomenon based on sensorimotor associations. The second dimension relates to more conscious, declarative and linguistic representations. Importantly, these two dimensions are then linked to the methodological perspective, where Mr. Majchrowicz reviews various methods used in the literature for assessing sense of agency. In line with the distinctions made in the phenomenological perspective, the methods can be grouped into explicit and implicit measures. The first are based on various types of questionnaires assessing participants’ subjective judgment of agency, while the latter are based on well-established paradigms, such as the intentional binding or sensory attenuation. The final part of the Introduction is related to mechanisms underlying sense of agency. Those are grouped in four paragraphs: complexity and integration of cues, temporality, learning and neural underpinnings. Complexity/integration of cues refers to information that the brain collects for the feeling or judgment of agency. Here, the key aspect is prediction of action outcomes and how precise this

Istituto Italiano di Tecnologia
Via Morego 30
16163 Genova Italy
Tel +39 010.71781

prediction can be. Mr. Majchrowicz discusses also other cues, such as social and emotional context, prior causal beliefs, action-effect contingency. Temporality refers to the aspects of prediction and inference, the first being related to the ability to anticipate sensory effects of actions while the second being related to postdictive processes such as outcome monitoring. The paragraph on learning is dedicated to an interesting and novel aspect considered in the context of sense of agency research, namely the interplay between feedback/reinforcement learning and sense of agency. According to the literature, the type of feedback (positive vs. negative) might influence sense of agency. In the final paragraph on neural correlates of sense of agency measures, Mr. Majchrowicz provides an overview of well-established imaging and electrophysiological effects observed in sense of agency literature.

Evaluation: The Introduction is very clear, well-written and well structured. It provides an excellent preview of the research focus of the thesis. It also provides extensive literature overview, although perhaps the reviewed literature on sensory attenuation could be slightly extended (by including, for example, very relevant work by Sonja Kotz and colleagues). Interestingly, it describes both philosophical and experimental perspectives on the topic of sense of agency. However, as the topic is extremely broad and widely covered in literature, the Introduction reads slightly too succinct and would perhaps benefit from a lengthier coverage of the multifaceted topic of sense of agency.

The chapters following the Introduction have already been published, and since they have gone through a rigorous peer-review process, I believe this speaks for itself, regarding the excellent quality of the content of the papers and requires no further evaluation. Therefore, in what follows, I will evaluate only the contribution of Mr. Majchrowicz to each of the papers in the overall context of his PhD thesis.

The first paper “Unexpected action outcomes produce enhanced temporal binding but diminished judgment of agency” by Majchrowicz, B. and Wierchoń, M., published in 2018, in *Consciousness and Cognition*, 65, 310-324, focuses on the aspect of temporal (and content) predictability of action effects. In three experiments, the authors manipulated predictability of a tone occurring after a key press. The studies used a classical intentional binding paradigm with the Libet clock. In the first experiment, participants were first exposed to an association phase where they learned (in the operant condition), the standard tone being an effect of their keypress. As in a typical intentional binding experiment, apart from the operant condition, participants took part also in a baseline tone and baseline action conditions. The intentional binding effect was operationalized in the form of differences in judgment errors between operant and baseline conditions. In addition to the intentional binding effect, participants made also explicit judgments of agency by providing ratings in a 5-item custom-made questionnaire. In the Experimental phase, the action effect (tone) could either be the same one as in the association phase (standard), or in a very small amount of trials, it could be a deviant, differing from the standard both in terms of temporal delay from the key press as well as the sound quality. The results showed that deviants produced larger intentional binding effect, relative to standards, which was an unexpected result against hypotheses based on previous literature. Interestingly, the explicit agency ratings followed the expected pattern, with most agency attributed to the standard outcome condition and less to deviant conditions. Experiment 2 tested whether the unexpected effect would replicate, and whether it would be modulated by various temporal delays. The results showed again the same pattern: replicating the “hyper” intentional binding effect (this time even stronger, and consistent across action judgment errors and tone judgment errors), and again a dissociation between the implicit and explicit measures.

The third experiment was designed to disentangle temporal and “content” predictability. In some conditions, the deviants differed from standards in terms of temporal delay (temporal predictability) and

in some others in terms of tone quality. The results showed that the novel “hyper”-intentional binding effect is observed only in case of temporal unpredictability. The authors discuss the results mainly in the context of pre-activation mechanism, although other interpretations of the unexpected hyper-intentional binding effect are also offered. The interesting result of this paper is also the dissociation between implicit and explicit measures.

Evaluation: The paper is methodologically solid and ambitious. Statistical analyses are advanced. The author contribution section specifies the Mr. Majchrowicz was responsible for all phases of the project, including funding acquisition, which is impressive for a PhD student. I think that Mr. Majchrowicz deserves a particular praise for the ambitious aspects of the paper, namely the advanced level of statistical analyses used (linear mixed regression models, Bayesian statistics), and for systematic approach to examining factors contributing to the effects of interest. This is particularly praiseworthy, given that this paper was published already in 2018, which means that the work has been conducted at early stages of the PhD project of Mr. Majchrowicz.

The second paper “Learning from informative losses boosts the sense of agency” by Majchrowicz et al. published in 2020, in *Quarterly Journal of Experimental Psychology*, 73, 2272-2289, is dedicated to the topic of relationship between sense of agency and learning. In two experiments, the authors examined how actions that produce losses influence sense of agency in subsequent trials, with sense of agency measured as intentional binding. The authors designed an experimental protocol combining the intentional binding paradigm with probabilistic reversal learning and task switching. In Experiment 1, the main question of interest was whether a previously observed phenomenon of “post-error agency boost” (PEAB) is task-specific or a generalized mechanism. For this reason, a task-switching component has been introduced. Interestingly, the two tasks differed with respect to the type of feedback that was given to participants (reward/loss communicated either via a picture of a coin or by means of a face with positive or negative emotion). Participants were asked to press a key which was followed by a sound and a feedback picture. The probability of winning/losing was set to 80/20. The results showed a PEAB effect independent of whether the feedback was signaled by means of a face or a coin picture. Interestingly, however, the effect was present only in task repetition trials, and not in task switch trials. On the other hand, the effect of current error (Current Error Agency Boost, CEAB) was observed for task switch trials, and mainly for the monetary reward feedback type. The authors interpreted the PEAB effect related to learning as task-specific, thereby indicating that the mechanism is present when people have the opportunity to adjust their behaviour in the process of learning. On the other hand, the CEAB was interpreted as a mechanism more related to processing of feedback valence, perhaps a precursor to the PEAB. In the second experiment, the authors examined the role of free choice in key pressing vs. instructed key press assignment. In addition, they measured the EEG signal and focused on two components commonly examined in the context of sense of agency, namely the P300 and the Feedback-Related Negativity (FRN). The task switching this time meant switching between free choice and instructed responses, rather than emotional vs. monetary depiction of outcome. The results showed PEAB effect for task repetition trials in the free choice condition and for task switch trials in the instructed condition. This suggests that a loss in a free-choice condition elicits boost in sense of agency on a subsequent trial, even when that subsequent trial does not involve free choice. The authors interpreted this effect as a mechanism for learning, when adaptation of behaviour is in principle possible. Interestingly, the FRN effect showed a pattern opposite to what would typically be observed, namely weaker FRN after losses as compared to winning trials. Finally, P300 showed higher amplitudes for previous losses, relative to wins. The authors interpreted the unexpected FRN effect as either being related

to predictability masking the outcome weighting, or to enhancement of FRN with reduction of sense of agency – an effect that has been found in literature before. P300 on the other hand was interpreted as unspecific mechanism signaling prediction error. Overall, the authors interpret the results in the context of the meaning of the PEAB effect as reflecting a mechanism that allows learning from mistakes in a specific task context, and prevents from learned helplessness.

Evaluation: the paper reports an extremely ambitious set of experiments, with complicated design and methods. The experiments have been conducted at UCL, which demonstrates that Mr. Majchrowicz made a research visit there to conduct the studies. This demonstrates high level of scientific maturity and independence of Mr. Majchrowicz. In addition, the complexity of the design and the analyses performed demonstrate methodological mastery of the PhD candidate. As with paper 1, Mr. Majchrowicz’s work on this paper deserves praise for ambition, methodological complexity and soundness. In addition, and perhaps most importantly, by combining experimental paradigms that classically belong to very different research traditions (task switching, intentional binding, probabilistic reversal learning), Mr. Majchrowicz showed an outstanding degree of creativity and original thinking in addressing novel theoretical questions and bridging gaps between different traditions.

The third, and final, paper of the thesis “Sensory attenuation of action outcomes of varying amplitude and valence” by Majchrowicz, B. and Wierchoń M. published in 2021 in *Consciousness and Cognition*, 87, 103058, focuses on the phenomenon of sensory attenuation. Sensory attenuation has been observed in the literature in the form of lower perceived intensity of a sensory signal which is a consequence of one’s own action, relative to externally generated sensory signal. This effect has been also commonly investigated with electrophysiological measures, where attenuated amplitude of an event-related-potential (ERP) of an EEG signal has been observed for self-generated vs. externally-generated sensory events. The authors addressed the sensory attenuation phenomenon in the context of two questions: the impact of intensity of the sensory signal, and the impact of its emotional valence. These two factors were chosen in order to test predictions of three different theoretical accounts related to sensory attenuation, namely the cancellation account (related to the forward model and efference copy of motor commands), pre-activation account (related to pre-activation of neural activity in sensory areas), and expected intensity account, which has different predictions related to attenuation/amplification, depending on the intensity of the sensory event. In Experiment 1, participants performed a task, in which their key presses were followed by a sound of two different intensities. Sensory attenuation was measured on a visual analogue scale. The authors did not find any evidence for sensory attenuation, only a weak evidence for sensory amplification. In Experiment 2, in addition to intensity of the two sounds, the authors added a factor of valence. Participants’ key presses were followed by human vocalization sounds with either positive or negative valence. In addition to the amplitude rating on a visual analogue scale, participants performed also valence rating and filled out explicit sense of agency questionnaire. The results showed again no evidence of sensory attenuation, and only a weak evidence for sensory amplification. For valence rating, quiet negative outcomes were rated more negative in the active, as compared to passive hearing condition. This was interpreted by the authors as indication of adaptive function of sense of agency in terms of distancing with respect to negative outcomes of one’s own actions. Finally, the explicit judgment of agency showed no significant effects related to the manipulation. Overall, the authors concluded that the sensory attenuation effect might not be as stable as is depicted in the literature, and that it might depend on various factors, emotional valence being one of them.

Evaluation: This is an interesting paper addressing timely question and exploring certain gaps in understanding factors influencing sensory attenuation. Methodologically, it is less complex than the second paper, but it is very solid, and theoretically well embedded. Mr. Majchrowicz's contribution to the paper was major, and demonstrates his scientific maturity through the comprehensive description of extensive theoretical background, and through a very balanced and comprehensive discussion of results that are difficult to interpret, as they are not in line with typically observed results. Once again, Mr. Majchrowicz demonstrated his excellent methodological and statistical skills through sophisticated analyses both through frequentist and Bayesian statistics.

Overall evaluation

The thesis presents a rich body of work, documented by three papers published in high quality and high-impact international journals. The contribution of Mr. Majchrowicz to this work is clear, and major. The work combines various methods (behavioral, EEG, subjective ratings) and various paradigms. In fact, some of the studies (Paper 2) have been conducted with experimental protocols combining various paradigms and experimental traditions. Statistical approaches are advanced, and for this Mr. Majchrowicz deserves particular praise. Finally, some studies have been conducted abroad, with funding acquired by Mr. Majchrowicz himself. This demonstrates a high level of scientific maturity and independence. It also demonstrates the ability to establish international collaborations and to work in various research environments.

The thesis itself is well written, putting together the three papers in one coherent narrative. Perhaps only one element that I have missed in the thesis is a theoretical chapter (either at the end of the Introduction or at the end of the thesis) which would embed, in a theoretical account, all results reported in the thesis. In particular, it would be important to reconcile within a theoretical framework the intriguing results observed in Paper 1 and Paper 3, namely enhanced intentional binding for unexpected outcomes (Paper 1) and sensory amplification, rather than attenuation (Paper 3). At the end of the thesis, I was left with an unanswered question: how can these results be reconciled with existing theoretical accounts of sense of agency (e.g., the comparator model or the pre-activation account)? Can any of the existing models or theoretical frameworks be extended to account for all the effects reported in the thesis? This approach is partially undertaken in discussion sections within Paper 1 and Paper 3, but the conclusions are rather negative stating that the reported results do not (fully) support existing models. However, it would be interesting to find in the thesis a proposal for a theoretical account (or an extension of the existing models) that could actually explain all the reported results. In the Introduction chapter, Mr. Majchrowicz mentions the "Bayesian cue integration" framework as a possible candidate framework, in Paper 1, the "Pre-activation" account seems to be the best candidate. It would be very interesting to read a more elaborate proposal how all observed effects (across all three papers) can be incorporated a potential candidate theoretical framework. This is, however, only a minor issue in the context of all the extensive work that has gone into the entire PhD project of Mr. Majchrowicz. Therefore, I can certainly evaluate the thesis as excellent, and wholeheartedly **recommend Mr. Majchrowicz to pass to the next stage of the PhD award process. I recommend the thesis to be evaluated with distinction.**

Sincerely,



Prof. Agnieszka Wykowska

Istituto Italiano di Tecnologia
Via Morego 30
16163 Genova Italy
Tel +39 010.71781