Emotional valence, sense of agency and responsibility: a study using intentional binding

*Christensen, J.F. [1], Yoshie, M.[2], Di Costa, S. [1], Haggard, P. [1]

[1] Institute of Cognitive Neuroscience, University College London (UCL)

[2] Human Informatics Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Japan

* Corresponding author: Julia F. Christensen; Institute of Cognitive Neuroscience, University College London (UCL); Alexandra House, 17 Queen Square, London WC1N 3AR; j.christensen@ucl.ac.uk

Abstract

We investigated how the emotional valence of an action outcome influences the experience of control, in an intentional binding experiment. Voluntary actions were followed by emotionally positive or negative human vocalisations, or by neutral tones. We used mental chronometry to measure a retrospective component of sense of agency (SoA), triggered by the occurrence of the action outcome, and a prospective component, driven by the expectation that the outcome will occur. Positive outcomes enhanced the retrospective component of SoA, but only when both occurrence and the valence of the outcome were unexpected. When the valence of outcomes was blocked –and therefore predictable– we found a prospective component of SoA when neutral tones were expected but did not actually occur. This prospective binding was absent, and reversed, for positive and negative expected outcomes. Emotional expectation counteracts the prospective component of SoA, suggesting a distancing effect.

Keywords: Sense of agency, Intentional Binding, Valence, Emotion, Retrospective, Prospective

Emotional valence, sense of agency and responsibility: a study using intentional binding

1. Introduction

Sense of agency is the experience of control over our own voluntary actions, and, through them, events in the outside world. Sense of agency is a core feature of normal mental life, at least in healthy adult humans, and a prerequisite of a well-functioning society. For example, legal systems assume that we will always *know* whether we have caused an event or not, because we *experience* our actions, and, at least partly, understand their consequences, at the time that we make them. Other things being equal, a defendant who has a sense of agency regarding the appropriate action is expected to plead *guilty*, and one who does not is expected to plead *not guilty* to the appropriate charge.

If the sense of agency perfectly tracked the facts of agency, there would never be doubt about who caused a particular outcome, and was thereby responsible. However, the sense of agency is limited for several reasons. First, sense of agency does not perfectly track the objective facts of agency. For example, a person may make an action and cause an outcome, but not realise they have done so. One reason for this gap between objective and subjective agency is limited cognitive capacity: people clearly cannot foresee all the consequences of their own actions. Here we focus on a different limitation, namely the strong biasing effect of affective valence on sense of agency. These effects are often considered under the label self-serving bias (Heider, 1958; Bandura, 1982; Bradley, 1978). People attribute positive actions and outcomes to themselves, while distancing themselves from bad actions and outcomes, notably by attributing them to others. Previous studies of such biases have used a framework of attribution judgement to consider these biases, often within the social context of praise and blame.

However, the relation between such social judgements of agency, and the primary experience of agency, remains unclear. This issue is important, since people often know from direct experience that they are responsible for an outcome, yet then explicitly attribute responsibility to another (e.g., "passing the buck", "only obeying orders"; Miller, 1947, 2009). Such psychological phenomena imply a basic, pre-attributional experience of agency, which is sensitive to valence. However, this experience, and its valence-sensitivity have proved difficult to measure scientifically, resulting in a knowledge gap in the literature on psychology of agency and responsibility.

Here we have used action binding (Haggard, Clark, & Kalogeras, 2002; Moore & Obhi, 2012) as an implicit measure of sense of agency. The perceived time of a voluntary self-generated action is shifted towards the action's outcome, as compared to a condition where the action is not followed by a sensory outcome. There is also a corresponding shift in the perceived time of the outcome back towards the action that caused it.

Yoshie & Haggard (2013) compared intentional binding for actions which predictably had either positive, negative or neutral sounds as their outcomes. Binding was reduced for negative action outcomes as compared to positive ones. This implies a stronger sense of agency over positive compared to negative events, consistent with the concept of a self-serving bias (Bandura, 1984), and/or a distancing from negative outcomes as a form of

reappraisal mechanism (Ochsner, Bunge, Cross, Gabrieli, 2002). Takahata et al. (2012) obtained similar results in the context of a gambling task. Different tones were first associated with either positive or negative outcomes of one's own action (gains or losses). Subsequent intentional binding measures showed less binding, implying reduced sense of agency, for tones that had previously been associated with losses, compared to rewards.

The experience of agency over an outcome could be based on predicting the outcome (e.g., I feel in control of my bicycle, because I predict that turning the handlebars will change my direction), or on inferring retrospectively the consequences of my own actions (e.g., I win money because I gambled. Although the stakes were 50-50 my winning makes me retrospectively attribute the winning to my skill in choosing) (Moore & Haggard, 2008). It remains unclear whether valence effects on sense of agency are primarily prospective, or primarily retrospective. In previous studies (Yoshie and Haggard, 2013), the valence of an action was entirely predictable, since participants generated positive, negative or neutral outcome sounds in separate blocks. In this situation, both prospective and retrospective components can contribute to sense of agency, but cannot be specifically disentangled. In Takahata et al. (2012), the valence of the outcome was randomised, and therefore unpredictable. In their results, outcome valence influenced sense of agency retrospectively. However, it remains unclear whether valence can also influence prospective sense of agency.

This issue has important implications. Purely retrospective valence effects imply a failure to feel responsible for actions with negative outcomes. This would be highly adaptive in ensuring well-being, but would have catastrophic implications for society. A well-functioning society, at least one similar to our own, presupposes that individuals avoid making actions with negative outcomes, even when these actions are superficially tempting. Presumably, individuals learn from previous experience the relation between actions and negative outcomes, and then use these experiences to prospectively guide future agency. A genuine, valence-sensitive experience of agency at the time of an action therefore plays an important role in minimising future harmful actions.

Moore and Haggard (2008) proposed an experimental design to distinguish the influence of prospective and retrospective mechanisms on sense of agency, based on manipulating outcome probabilities. The probability of producing an outcome is set to 50% in one block of the experiment, and to 75% in another (Moore & Haggard, 2008; Voss, et al., 2010). An estimate of the retrospective component of sense of agency is obtained by comparing the binding of action towards tones on those trials in the 50% block where a tone does occur, with those trials where it does not. An estimate of the prospective component is obtained by comparing the binding of actions towards tones on trials where tones do not in fact occur, but are more likely (i.e., 75% block), compared to less likely (i.e., 50% block). Prediction should clearly be stronger in the 75% block than in the 50% block, although it is not total in the former, nor absent in the latter – thus the design affords a partial estimate of the prospective component, rather than a perfect measure. Nevertheless, comparing estimates for prospective or retrospective components between groups (Voss et al., 2010), or between different conditions may clarify whether a particular factor influences prospective or reconstructive components of sense of agency.

We investigated how the valence of an action outcome would influence the retrospective and prospective components of sense of agency. The probability of producing an action

outcome was either highly likely (75%) or moderately likely (50%), and retrospective and prospective components of action binding were calculated. In Group 1, outcome *valence* was *unpredictable*, because positive, negative and neutral sounds were randomised. This means that for this group, the valence of the outcome, if it occurred, was experienced anew on each trial. This allowed us to assess whether the outcome valence of an action modulates sense of agency *retrospectively*. In Group 2, the same outcome sounds were blocked, so that the *valence* of the sound was always *predictable*, though its *occurrence* could be more or less predictable, as before (50% or 75%). With this we sought to establish whether the prospective component of sense of agency would emerge if the valence of the outcome, as well as its occurrence, were predictable.

Our general research question was whether outcome *valence* influences the prospective or the retrospective part of sense of agency, and whether this influence depended on *valence* itself being predictable.

2. Method

2.1. Participants

A total of 56 right-handed volunteers (30 male) participated in the experiment in exchange for a small time reimbursement (£7.50/h). They were randomly assigned to the two conditions (*randomised,* unpredictable emotion trials, *vs*. *blocked,* predictable emotion trials); Group 1 (28; 14 male; mean age = 23.54, SD = 4.53) and Group 2 (28; 16 male; mean age = 24.25, SD = 5.49). Emotion predictability was used as a between rather than within subjects factor, largely because a within subjects approach would have shown an order effect (once emotions are predicted, they may continue to be predicted, even after they become unpredictable). All participants were of European Linguistic background to match the actors of the vocal sound stimuli.

To determine the sample size of the two groups, we performed two power calculations, using *GPower 3.1.* (Faul, Erdfelder, Lang, & Buchner, 2007) for two contrasts of interest. First we determined the sample size necessary to identify a significant effect of emotion within a single group (effect size = .80; alpha = .05; power = .90). This gave a suggested sample size of 15. In addition, we also planned to compare the group receiving blocked emotion outcomes with the group receiving randomized emotion outcomes. For this, we performed an additional power calculation for an independent samples t-test (effect size = .80; alpha = .05; power = .90). This gave a suggested sample size of 28 per group. The total sample size was fixed based on this second power calculation.

2.2. Procedure

An Intentional Binding task was used as an implicit measure of sense of agency (Haggard, et al., 2002; Libet, et al., 1983). Participants were seated in front of a computer screen with an external silent SODIAL(R) Flexible Foldable USB keyboard in front of them to provide their responses on. At each trial, participants were instructed to press a key [ENTER] on this keyboard at the time of their choosing (= the moment they felt the urge to
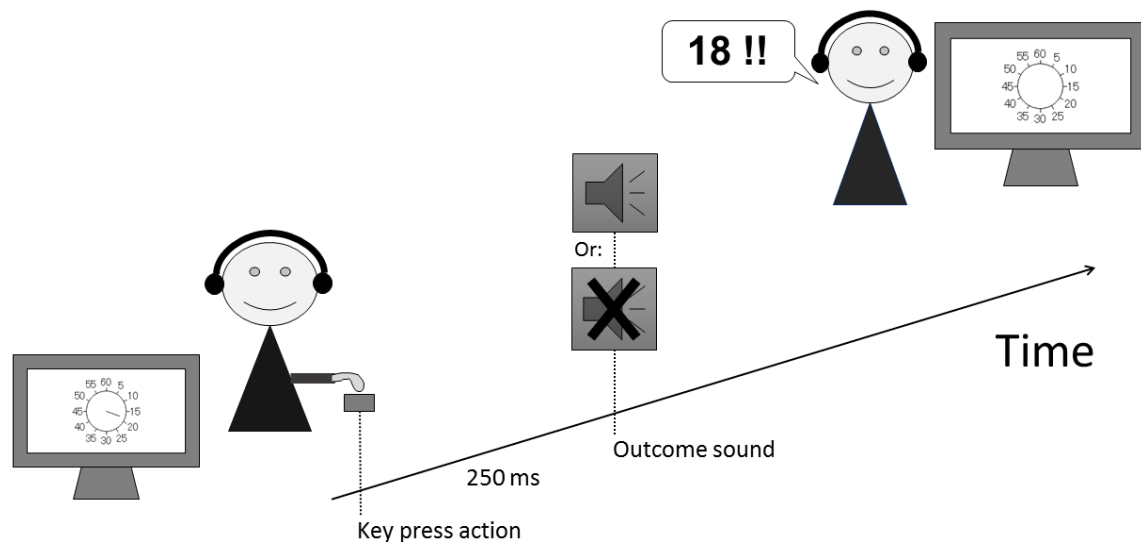
5

do so) while they looked at the centre of a clock which was displayed on the screen with a continuously rotating clock hand. The clock was ~2x2 cm large and the clock hand 12mm, rotating about a clock face with the common numbers 5….10….15 etc., at a rate of 2,560 milliseconds (ms). After the key press, after a fixed duration of 250 ms a tone occurred. This brief interval between action and tone gave participants the impression of causing the tone. The clock hand continued rotating for a random amount of time, then stopped. At this time, participants were prompted to state verbally where the clock hand was in the moment they pressed the key. The experimenter entered the number on a different keyboard and launched the next trial.

In the classical action binding procedure, the binding measure is obtained by calculating the difference in estimation accuracy between the following two conditions: first, the participant performs this task in a baseline condition where no tone occurs. In the operant condition, they generate the tone by their own voluntary action. Their judgment "shifts" towards the tone, signalling that the outcome of their action (the tone) modulated their temporal estimation of the time of the action. This difference between baseline and operant conditions is defined as "action binding", and serves as a proxy for sense of agency.

In order to give the action outcomes different affective significance, three types of emotional sounds were selected, following previous empirical work (Yoshie & Haggard, 2013). Eight emotional sounds were selected from an extensively-normed database of nonverbal emotional vocal sounds (Sauter, Eisner, Calder, & Scott, 2010). There were four sounds with positive valence (two sounds of laughter and two of achievement), and four of negative valence (two of disgust and two of fear). In addition, four neutral "beep"-tones were generated to be used as a neutral comparison condition. This resulted in a total of 12 outcome sounds, of three different valence significance. Subjective ratings from an independent group of participants in a previously published study confirmed this classification of the sounds as positive and negative, and also verified matching for pitch and duration; see supplementary material of Yoshie & Haggard, 2013).

Prospective and retrospective processes in sense of agency can be investigated using the intentional binding paradigm by varying the outcome probability across two conditions, one 50% and one 75% condition (Moore & Haggard, 2008; Voss, et al., 2010). The same procedure was followed here. See figure 1 for an illustration of the basic trial structure.

Figure 1. Trial structure.

Participants performed a total of 240 trials. Both groups started with an action baseline block (48 trials). Group 1 (*randomised* unpredictable emotion group) subsequently went through 4 blocks of 48 trials. In 2 of the blocks the sound occurrence predictability was 50% (in only 50% of the trials a sound actually occurred) and in another 2 blocks sound occurrence predictability was 75% (in 75% of the trials a sound occurred). In the 50% predictability condition participants had a total of 16 negative, 16 positive and 16 neutral sound outcomes (=48), and a further 48 trials with no sound outcome (randomised across 2 blocks, with equal numbers of each trial type in each block). In the 75% predictability condition participants had 24 stimuli of each valence (72 trials) plus 24 trials with no sound outcome (randomised across 2 blocks, equal number of trials of each in each block). This means a total of 96 in each of the 2 conditions (=192), plus 48 from the baseline condition (=240 in total).

There were two groups of participants. Group 2 (the predictable *blocked* emotion group) received the same number of stimuli of each kind as the group in Group 1. To allow the blocked paradigm, however, the stimuli were arranged as follows: participants in Group 2 performed 3 shorter blocks of 32 stimuli for each of the two outcome probabilities (50% and 75%), 6 in total; 2 for each valence (negative, neutral, positive). In each of these blocks, there were 16 stimuli of the respective valences and 16 without any sound (50% condition), or 24 of each valence and 8 without sound outcome (75% condition). This resulted in 32 trials in 6 blocks (=192) plus the 48 from the baseline (=240 in total). In both experiments, blocks were counterbalanced between participants but participants always performed either the 50% or the 75%. Participants always performed the baseline condition first, then followed by the blocks of the experiment, in counterbalanced order. See figure 2 for an illustration of the procedure and table 1 for the design.

Figure 2. Schematic representation of the procedure. (A) Voluntary actions are followed by an outcome sound (beep) 250ms later. This produces a shift in the perceived time of the action (dotted arrow). (B) In one block, the probability of a sound given an action is 50%. Binding toward outcome is compared between trials with and without outcome sound. This identifies the retrospective component of action binding. (C) In another block, the probability

given an action is 75%. Comparing action binding on trials across the two probability blocks where no outcomes occur, reveals how outcome binding varies with the degree of prediction.
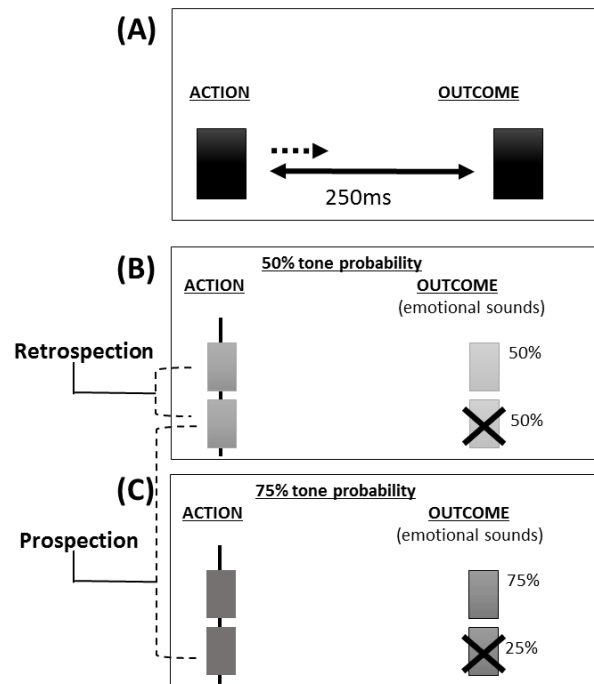

Table 1

The experimental design was based on the factorial combination of two factors. Outcome occurrence was a within-subjects factor, referring to whether the outcome was more (75%) or less (50%) likely. Outcome valence was a between-subjects factor, and was either unpredictable (Group 1) or predictable (Group 2).

	OUTCOME OCCURENCE	
	Less Predictable	*More Predictable*
VALENCE	*Probability of sound given an action = 50%*	*Probability of sound given an action = 75%*
UNPREDICTABLE (GROUP 1)	No Sound 50% / Sound 50%	No Sound 25% / Sound 75%
PREDICTABLE (GROUP 2)	No Sound 50% / Sound 50%	No Sound 25% / Sound 75%

To ensure participants would not ignore the sounds' emotional significance, participants were informed that a few randomly interleaved "extra questions" would appear throughout the experiment, asking them to state whether the sound they just heard was positive or negative (this question was never asked after a neutral beep). Participants were told they would receive a bonus (25 pence) for each correct answer in addition to their normal payment. There were 10 such trials throughout each experiment. Participants in both groups generally judged between 8 and 10 of these trials correctly (average: 9.80; SD 1.10). One outlier scored only 2/10.

2.3. Statistical analyses

We calculated a retrospective and a prospective component of action binding, following previous studies (Moore & Haggard, 2008), and we analysed these components separately. The *retrospective component* of action binding was calculated as the difference in action binding in the 50% occurrence probability condition between trials where there *was* an outcome sound and trials where there was *no* outcome sound. For Experiment 1 the no sound trials were averaged across all three valences, as these trials were presented at random and could not be allocated to any specific valence. For Experiment 2, each valence was presented in a separate block, and we therefore used the no sound trials from the relevant block to calculate the retrospective component: (positive outcome)-(no outcome in the positive block), (negative outcome)-(no outcome in the negative block), and (neutral outcome)-(no outcome in the neutral block).

The *prospective component* was calculated by subtracting the average action binding in the *no sound* trials of the 50% condition from the average action binding effect in the *no sound* trials of the 75% condition. The prospective component therefore captures the extent to which action binding depends on the probability of the outcome, given the action.

A Repeated Measures design was used with separate blocks of judgments of action with outcome sound *valence* as within-subjects factor (negative, neutral or positive) and *predictability* of outcome valence as between-subjects factor (unpredictable; Group 1, *vs.* predictable; Group 2). Specifically, in Group 1 emotional outcomes were fully randomised to create circumstances of fully unpredictable emotional outcomes. In Group 2 the emotional outcomes were blocked into separate blocks of predictably negative, neutral or positive outcomes.

As effect sizes we report partial eta ($\eta_p{}^2$), where .01 is considered a small effect size, .06 a medium effect and .14 a large effect, and Cohen's d for t-tests (Cohen, 1988).

3. Results

The data are shown in table 2.

Table 2
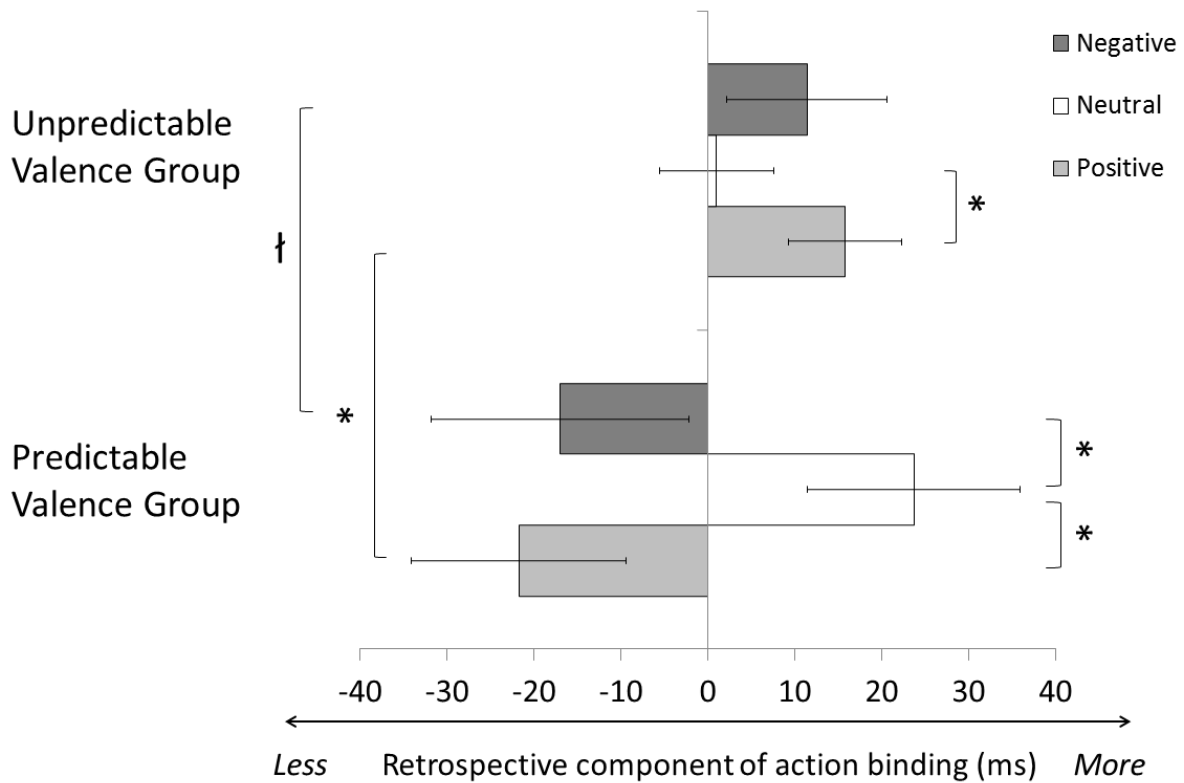Results of experiments 1 and 2. Values are rounded to the nearest ms.

CONDITIONS	Outcome	Group 1 (unpredictable emotion trials)		Group 2 (predictable emotion trials)	
		Mean	*SD*	Mean	*SD*
Baseline	-	-36	57	7	58
Probability of Sound 50%					
	No outcome (average)	28	69	0	60
	No outcome –negative	*N/A[1]*	*N/A*	15	83
	No outcome –neutral	*N/A*	*N/A*	-6	56
	No outcome –positive	*N/A*	*N/A*	13	52
	Negative outcome	40	72	-2	70
	Neutral outcome	29	71	10	56
	Positive outcome	44	68	-9	70
Probability of Sound 75%					
	No outcome (average)	28	71	1	53
	No outcome –negative	*N/A*	*N/A*	-1	62
	No outcome –neutral	*N/A*	*N/A*	12	55
	No outcome –positive	*N/A*	*N/A*	-1	58
	Negative outcome	35	73	14	65
	Neutral outcome	28	62	18	63
	Positive outcome	26	76	1	64
Retrospective component					
	Negative outcome	11	36	-17	78
	Neutral outcome	1	35	24	65
	Positive outcome	16	26	-22	65
Prospective component					
	Average over outcomes	0	29	0	41
	Negative outcome	*N/A*	*N/A*	-16	72
	Neutral outcome	*N/A*	*N/A*	18	46
	Positive outcome	*N/A*	*N/A*	-13	55

[1] These values cannot be calculated for Group 1 because the emotional sounds were randomised across trials and thus the "no sound" trials cannot be attributed to any specific emotional valence condition.

Our analyses focussed on the retrospective and prospective components, rather than the individual cell means. The results are shown in figure 3. The *Retrospective component* was analysed using a mixed Repeated Measures ANOVA with between-subjects factor of

predictability of outcome valence (unpredictable, vs. predictable), and a within-subjects factor of outcome sound valence (negative, neutral or positive). We found no significant effect of either factor (predictability: $F(2,53) = 2.425$, $p = .125$, $\eta^2 = .043$; outcome valence: $F(2,53) = 1.638$, $p = .199$, $\eta^2 = .029$), but a significant interaction between these factors ($F(2,53) = 5.581$, $p = .005$, $\eta^2 = .094$). Simple effects tests were used to explore this interaction. First, we compared the groups receiving predictable and unpredictable valence of outcome sounds, for each outcome sound individually. There was no difference between these groups for the neutral outcome sound ($t(54) = -1.631$, $p = .109$, ns). However, the groups did differ for the negative ($t(54) = 1.743$, $p = .087$, $d = .46$) and for positive sounds ($t(54) = 2.835$, $p = .006$, $d = .77$). Action binding was stronger when these sounds were unpredictable (Group 1), compared to predictable (Group 2). In addition, we also used simple effects to explore differences between the different valences of outcome sounds within each group. In group 1 for whom outcome valence was unpredictable, we found more binding for positive than for neutral sounds ($t(27) = 2.069$, $p = .048$, $d = .48$). No such difference was found between positive and negative sounds ($t(27) = -.588$, $p = .562$, ns), nor between negative and neutral sounds ($t(27) = 1.407$, $p = .171$, ns). In group 2, for whom outcome valence was predictable, we found a significant difference in action binding between negative and neutral sounds ($t(27) = -2.388$, $p = .024$, $d = .57$), and between positive and neutral sounds ($t(27) = 2.779$, $p = .010$, $d = .70$), though in the opposite direction as compared to group 1. Group 2 showed less action binding for negative and positive outcomes, as compared to neutral outcomes.
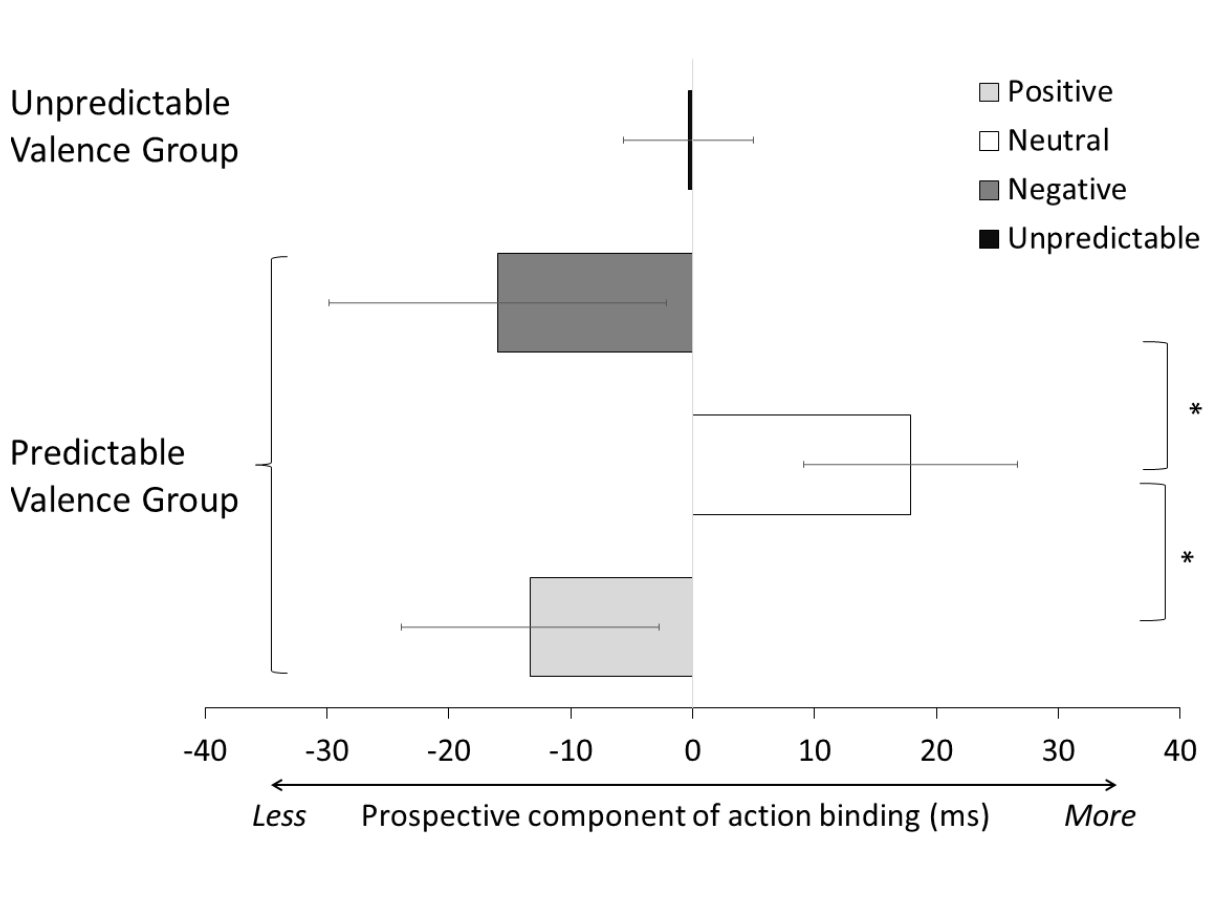
Figure 3. Retrospective component of sense of agency. See text for explanation. Error bars represent S.E.M. * *p* < .05. † p < .10

The *prospective component* could not be analysed factorially. In the condition where valence of outcome sounds was blocked (Group 2), we could compare the prospective components for each valence condition. However, in the condition where valence of outcome sounds was randomised (Group 1) and therefore unpredictable, we cannot assign individual no-sound trials to any particular valence condition, and could therefore not calculate a prospective component for each valence. We therefore performed two separate analyses. First, we used a between-subjects t-test to compare groups receiving unpredictable or predictable valence of outcome sounds, averaging over the different outcome sound valences in the latter case. There was no significant difference between the groups ($t(54) = -.067$, $p = .947$, *ns*). Second, for the predictable valence group (Group 2) only, we used a one-way ANOVA to compare how valence influenced the prospective component. This showed a significant overall effect of outcome sound valence ($F(2,27) = 3.195$, $p = .049$, $\eta^2 = .106$. Follow-up t-tests showed significant differences between the prospective component when a negative sound might be predicted compared to when a neutral sounds might be predicted ($t(27) = -2.424$, $p = .022$, $d = .75$), and between the prospective component when a positive sound might be predicted compared to when a neutral sound might be predicted ($t(27) = -2.359$, $p = .026$, $d = .62$). The difference between negative and positive conditions was not significant ($t(27) = -.157$, $p = .876$, *ns*). No correction for multiple comparisons was

performed: when omnibus ANOVA shows a difference between three conditions, subsequent pairwise comparisons do not use additional degrees of freedom (Cardinal & Aitken, 2006).

Figure 4. Prospective component of sense of agency. See text for explanation. Error bars represent S.E.M. * $p < .05$.

4. Discussion

We have investigated the relation between the valence of an outcome and the sense of agency in a laboratory experiment with healthy adult volunteers. Previous work on agency-valence relations has traditionally been dominated by the pervasive concept of self-serving bias. This implies enhanced sense of agency for positive outcomes, and a distancing, or reduced sense of agency for negative outcomes (Bandura, 1984). Crucially, our method combined a number of innovations relative to previous work: we used an implicit proxy measure of sense of agency, we distinguished between prospective and retrospective components of sense of agency, and we used naturalistic human vocalisations as action outcomes, ensuring a basic affective valence because of the emotional nature of these sounds.

Our study generated three main findings. First, unexpectedly positive outcomes enhanced the retrospective sense of agency. Second, when the valence of an outcome was expected, this retrospective effect was abolished, and in fact reversed –for both positive and negative outcomes. Third, the prospective component of sense of agency was enhanced when a neutral outcome was expected, compared to when a positive or negative outcome was expected.

13

Conversely, when there was no expectation regarding the valence of an outcome, the prospective component of sense of agency was reduced.

In principle, the negative scores found for prospective and retrospective components of action binding in group 1 and 2 could be due to some process other than affective modulation of sense of agency by action outcome. For example, in group 2, the predictable emotional valence of each action might favour learning time perception. However, participants never get feedback, so the opportunities for learning are limited. Therefore we disregard this interpretation of the current results.

Our results may reflect two qualitatively different interactions between the neurocognitive systems for emotion and for action. The first of these might be called *agentic serendipity*. When outcomes were (a) relatively unlikely, (b) of unpredictable valence, and (c) happened to be positive, we found a significant increase in the binding of an action towards the sound. We interpret this as a retrospective boost of sense of agency, consistent with self-serving bias (Heider, 1958; Bandura, 1984). Importantly, the conditions for this effect occurring are relatively restrictive: the effect was absent for unexpectedly negative sounds, ruling out an effect of mere salience. Further, the effect was absent in a group of participants for whom the valence of the sound was predictable. Thus, we confirmed a self-serving bias in sense of agency, and showed that it operates retrospectively, only once the positive valence of the outcome is known. These findings are consistent with an inferential mechanism, which attributes "nice surprise" effects to one's own voluntary agency. Interestingly, we found no comparable retrospective distancing effect. That is, unexpectedly negative outcomes were not associated with a reduced sense of agency. This may be reassuring in social and legal contexts: individuals are expected to accept responsibility for their actions. Even when an action has an unforeseen negative outcome, as in manslaughter, society still attributes the action to the agent. Our result suggests that this social attribution is consistent with the agent's low-level perceptual experience of their action.

Our concept of retrospective agency was based on changes in intentional binding when an outcome unpredictably occurred after an action, compared to when it did not. Importantly, however, we found that ability to predict the outcome's *valence* had a clear effect, over and above the ability to predict its *occurrence*. Group 2 of our study experienced each outcome valence in a separate block, and could therefore predict in advance the valence of an outcome, if one were to occur. In this group, the actual occurrence of a sound could be more or less predictable, according to the block, but the valence of the sound, if it occurred, was always predictable. Emotional valence of an unpredictable sound reduced the sense of agency, both when the sound was positive and when it was negative, as compared to when it was neutral. That is, the retrospective component of action binding was reversed when a predictably positive or predictably negative vocalization occurred. Because this effect was equivalent in positive and negative outcome blocks, it remains unclear whether it is based on valence, or merely on salience. The retrospective component of agency has often been associated with poor understanding of the relation between one's actions and their outcomes, with 'leaping to conclusions' about agency (Voss et al., 2010), and with uncertain contexts. Our results suggest these processes are reduced, rather than increased, by a stable emotionally-valenced context, whether positive or negative. When the valence of an outcome can be predicted, the tendency for emotional events to rapidly restructure experience is

14

abolished, and even reversed. Predictable, emotionally-valenced outcomes may produce a general emotional context, by a process similar to mood induction. While exogenous, bottom-up positive affect can augment retrospective inferences underlying sense of agency, the ability to predict emotional outcomes appears to decouple sense of agency from the influence of actual sensory evidence about outcomes, via a kind of emotional distancing. Thus, the agentic serendipity effect is cancelled when valence is predictable. This result may be related to altered sense of agency in affective pathologies such as depression and euphoria (Alloy & Abrahamson, 1979; Alloy, 1988, Msetfi, Murphy, and Kornbrot, 2012; The Diagnostic and Statistical Manual of Mental Disorders V, 2013).

The second link between agency and valence might be called 'cold prediction'. The prospective component of sense of agency was identified as an increase in binding of actions towards outcomes when outcome occurrence is more rather than less likely. We found no evidence for a prospective component when the valence of outcomes was unpredictable. When outcome valence was predictable we found a significant prospective component of binding only for neutral outcomes. In fact, greater predictability of negative or positive outcomes leads to less binding than lower predictability. This result suggests that sense of agency involves both a cold, prospective component and a warm, retrospective one. For both components, the negative-positive comparison is never significant, so the mechanisms seem to be sensitive to valence in general, that is, unsigned valence rather than signed valence. We recognise that unsigned valence and salience may be confounded, because positive and negative vocalisations may both be more salient than neutral sounds. Briefly, the prospective mechanism seems to be a cold and rational one: it is reduced, and even reversed when the affective valence can be predicted. Affect and cold prediction are thus mutually antagonistic. Conversely, in situations of low statistical predictability, the sense of agency appears to be driven by a postdictive mechanism based on positive affective surprise. This latter mechanism requires that both the occurrence and the valence of an action outcome be surprising: the effect was abolished when occurrence was unpredictable, but valence was predictable.

The legal concept of responsibility for action can be linked to specific views about the roles of volition and cognition in psychology of action. In particular, the law assumes that individuals have voluntary control over their actions, and that they understand the relation between their action and the consequences of those actions. For example, the M'Naghten rules (case 1843-60 All ER Rep 229) famously ask whether an individual knows "the nature and quality of the act" at the time of their action. At first sight, a hypothetical individual who simply lacks any sense of agency could not be found guilty under these rules (though they could instead be found 'criminally insane'). We have described prospective and retrospective aspects of sense agency. According to these concepts, an agent might *not know the nature and quality of their act* if they have lack an advance representation of their goal with which to guide their action control, and also lack any experience of linkage between their action and the actual outcome.

At the same time, the law frequently deals with situations where intentional action, emotional state, and outcome valence are co-occurring and important factors. Thus, any influences of affective valence on sense of agency may be relevant for legal concepts of responsibility. In *loss of control* defences (Sec. 54 Coroners and Justice Act 2009), an

ongoing, extreme emotional state is invoked to explain an agent's escape from normal voluntary control. In particular, loss of control may potentially explain why people perform actions with clearly negative outcomes. Our data showed that the prospective component of sense of agency is lost when an outcome is predictably negative, or predictably positive. We speculated that regular exposure to negative outcomes could cause a kind of negative mood induction, like a very diluted version of the depressive realism effect (Alloy & Abramson, 1979, 1988). In extreme form, these co-occurring elements may provide a psychological mechanism relevant to the loss of control defence. That is, regular extreme negative experience, combined with reduced ability to predict and register a potential negative outcome of one's own action could constitute a departure from normal sense of agency. Our key result is that affective context may change the experience of *the nature and quality of the act*. In particular, recurrent and predictable negative experiences appear to reduce the prospective sense of agency. The prospective component of sense of agency is thought to arise at the time of action, since it does not depend on actual outcome occurrence (Moore & Haggard, 2008). It may depend on the same brain processes that generate intentional actions (Moore, Ruge, Wenke, Rothwell, Haggard, 2010). To this extent, our study may clarify why, in criminal contexts, the normal cognitive mechanisms of action control seem to be replaced by emotionally-charged drives to action that lack the normal experiences of agency and responsibility. The existence of such a brain mechanism clearly does not excuse wrong doing. However, it does suggest that a feeling of reduced responsibility could potentially reflect an agent's actual experience during emotionally-charged action. Reduced responsibility could correspond to a fact of human psychology, rather than a hopeful story to avoid punishment. Clearly, these considerations do not change the way justice should be done, but rather highlight features of human psychology that justice sometimes may, or may not, wish to take into account.

Our study has some obvious limitations. First, it is based on an implicit proxy marker for sense of agency. Using implicit measure gives several advantages relative to asking participants to report sense of agency directly – these have been reviewed elsewhere (Jensen, Overgaard, di Costa & Haggard, 2015). On the other hand, intentional binding measures have been criticised. For example, Dewey and Knoblich (2014) noted poor correlations between intentional binding and explicit agency measures, and therefore suggested that these measures did not measure the same thing. Second, while implicit judgements may be useful in scientific investigations of cognitive mechanisms, their relevance for social concepts of responsibility is less clear. For example, the law deals largely with explicit self-reports about actions. Third, some of the effects in our study are relatively small, in some cases only just achieving the boundary of statistical significance. Replication in an additional sample would therefore be valuable. Fourth, the results might generalise poorly to real-life situations outside of the experimental laboratory. Finally, our emotional vocalisation stimuli may have been imperfectly designed. We could not readily match positive and negative emotional vocalisations for strength and recognisability. Thus, positive and negative valences may not have been equally convincing or successful in manipulating affect. Negative emotional stimuli tend to be relatively ineffective in mood induction experiments (e.g., Davis et al., 2011), and may induce the opposite states, i.e., laughter (Levenson, 2014). More naturalistic stimuli or alternative mood induction methods especially for negative emotions, may support

a stronger claim about the link between laboratory measures of sense of agency and personal responsibility for action.

5. Acknowledgements

6. References

Alloy, L. B., and Abramson, L. Y. (1988). Depressive realism: four theoretical perspectives.

Alloy, L. B., Abramson, L. Y. (1979). Judgment of contingency in depressed and non-depressed students: Sadder but wiser? Journal of Experimental Psychology: General, 108(4), 441-485

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders: DSM-5 (5th ed.)

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist,* 37, 122-147.

Bandura, A. (2002). Social cognitive theory: an agentic perspective. *Annual Review of Psychology, 52*, 1-26.

Bradley, G. W. (1978). Self-serving biases in the attribution process: A re-examination of the fact or fiction question. *Journal of Personality and Social Psychology, 36*(1), 56-71.

Cardinal, R. N., Aitken, M.R.F. (2006). ANOVA for the behavioural sciences researcher. Lawrence Erlbaum Associates, New Jersey, USA.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Coroners and Justice Act 2009. Sec. 54

Dewey, J. A., Knoblich G. (2014). Do implicit and explicit measures of the sense of agency measure the same thing? *PLoS One, 16;9*(10). doi: 10.1371/journal.pone.0110118. eCollection 2014.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 11*, 3656-3666.

Heider, F. (1958). The Psychology of Interpersonal Relations. New York: Wiley. Jones, E. E., D. E. Kannouse, H. H. Kelley, R. E. Nisbett, S. Valins, and B. Weiner, Eds. (1972). Attribution: Perceiving the Causes of Behavior. Morristown, NJ: General Learning Press.

Jensen, M., Di Costa, S., Haggard, P. (2015). Intentional binding: a measure of agency. In M. Overgaard (Ed.), *Behavioral Methods in Consciousness Research* (pp. 145-157). Oxford: Oxford University Press.

Levenson, R. W. (2014). The Autonomic Nervous System and Emotion. *Emotion Review, 6*(2), 100-112.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): the unconscious initiation of a freely voluntary act. *Brain, 106*, 623-642.

M'Naghten's Case [1843-60] All ER Rep 229.

Miller, A. (1947, 2009). All my sons. Penguin Classics: New York.

Moore, J. W., Haggard, P. (2008). Awareness of action: Inference and prediction. *Consciousness and cognition, 17*(1), pp. 136-44. ISSN 1090-2376

Moore, J. W., Ruge, D., Wenke, D., Rothwell, J., Haggard, P. (2010). Disrupting the experience of control in the human brain: pre-supplementary motor area contributes to the Sense of agency. Proce*edings. Biological sciences / The Royal Society, 277*(1693), pp. 2503-9. ISSN 1471-2954

Moore, J.W., & Obhi, S. (2012). Intentional binding and the sense of agency: A review. *Consciousness and Cognition, 21*(1), 546–561. doi:10.1016/j.concog.2011.12.002

Msetfi, R., M., Murphy, R. A., Kornbrot, D. E. (2012). Dysphoric Mood States are Related to Sensitivity to Temporal Changes in Contingency. Frontiers in Psychology, 3(368).

Ochsner, K. N., Bunge, S. A., Gross, J. J., Gabrieli, J. D. E. (2002). Rethinking feelings: an fMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience, 14*(8), 1215-1229.

Takahata, K., Takahashi, H., Maeda, T., Umeda, S., Suhara, T., Mimura, M., Kato, M. (2012). It's not my fault: Postdictive modulation of intentional binding by monetary gains and losses. Plos One. 7(12): e53421. doi:10.1371/journal.pone.0053421

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology (Hove), 63*, 2251-2272.

Voss, M., Moore, J., Hauser, M., Galligat, J., Heinz, A., & Haggard, P. (2010). Altered Awareness of action in schizophrenia: a specific deficit in predicting action consequences. *Brain, 133*, 3104-3112.

Yoshie, M., & Haggard, P. (2013). Negative Emotional Outcomes Attenuate sense of agency over Voluntary Actions. *Current Biology, 23*, 2028-2032.