© 2024 The Institute of Mind and Behavior, Inc.
The Journal of Mind and Behavior
Winter 2024, Volume 45, Number 1
Pages 1–28
ISSN 0271–0137

What Can the Side-Effect Effect Teach Us about How Intentionality Is Attributed to Individuals with a Psychiatric Disorder?

Christopher Papadopoulos

University of New South Wales (UNSW)

The side-effect effect (SEE) is the phenomenon whereby intentionality is more likely to be attributed to agents who bring about negatively valenced as opposed to positively valenced side-effects. The primary aim of the present series of studies was to examine whether the SEE would remain robust for judgments involving moral agents with a psychiatric disorder. A secondary aim was to provide a test of competing theoretical explanations of the SEE including the rational scientist model (Uttich and Lombrozo, 2010) and intuitive moralist accounts (Knobe, 2003a). This series of studies used psychiatric diagnostic labels to manipulate the norms participants applied to agents when judging intentionality. Intentionality ratings remained insensitive to normative information when agents were described as having a psychiatric disorder (Experiment 1a), when agents were also described as having an organic medical disorder that affected their behaviour (Experiment 1b), when the psychiatric norms were explicitly stipulated (Experiment 1c), and when participants with relative expertise in psychiatric norms were tested (Experiment 2). Taken together, the findings of this series of studies were more consistent with intuitive moralist accounts than the rational scientist model. Importantly, these studies extend the SEE to a novel paradigm and provide a demonstration of the robustness of the effect in the context of psychiatric diagnoses and the judgments of individuals with relative clinical expertise.

Keywords: side-effect effect, psychiatric disorder, bias

In a seminal study by Joshua Knobe (2003a) key issues pertaining to how people ascribe intentionality in realistic and morally charged situations were examined. Specifically, Knobe (2003a) designed scenarios involving a CEO whose primary aim was to achieve profit, and as a foreseen but unintended consequence (i.e., a side-effect) of doing so, impacted upon the environment. Crucially, the side-effect was either morally negative (the environment was harmed) or positive

This work was supported by an Australian Postgraduate Award to the author. Correspondence concerning this article should be addressed to Christopher Papadopoulos, Ph.D., School of Psychology, University of New South Wales, Sydney, NSW 2052, Australia. Email: education-eternal@outlook.com

(the environment was helped). When participants were asked whether the CEO intentionally harmed (harm condition) the environment most agreed (82%) whereas only a minority agreed that the CEO intentionally helped (help condition) the environment (23%). This asymmetry in intentionality ascriptions is referred to as the side-effect effect (SEE) and has been replicated widely and across a variety of paradigms and populations (e.g., Cushman and Mele, 2008; Knobe and Burra, 2006; Leslie, Knobe, and Cohen, 2006).

Understanding whether intentionality judgements are influenced by moral considerations is of major significance to attributions of responsibility to individuals with a psychiatric disorder, and therefore the SEE can potentially teach us about how responsibility is attributed to people with a psychiatric disorder. If an individual with a psychiatric disorder brings about some morally negative outcome, the SEE would suggest that responsibility for these actions would be attributed to the moral agent to the same or a similar degree as an agent without a psychiatric disorder because the moral content of the action (side-effect) remains unchanged by an agent's psychiatric status. However, this view is challenged by different perspectives which suggest that attributions of responsibility are mitigated for individuals with a psychiatric disorder.

A Psycho-Legal Perspective

There is scope within legal theory for holding people who suffer from psychiatric disorders to a different standard of responsibility for their actions than that applied to the larger population. This is based upon the assumption that certain psychological conditions can sometimes result in impaired judgement and/or impulse control (Diagnostic and Statistical Manual of Mental Disorders, fifth edition, DSM – 5; American Psychiatric Association, 2013). Perceived responsibility for the commission of illegal or immoral acts can be mitigated if an agent's psychological condition is judged to affect their ability to behave appropriately as reflected in the Anglo–American legal system (e.g., Australia, Canada, United Kingdom, United States, and Sweden) in the defences of insanity and diminished responsibility (see Anckarsäter, Radovic, Svennerlind, Höglund, and Radovic, 2009; Buchanan and Zonana, 2009; Gold, 2011; Mitchell, 1999 for discussion).

In cases where a crime has been proven, the insanity defence involves a full acquittal of all criminal charges on the basis of significant psychological incapacity (e.g., the defendant was diagnosed with schizophrenia and was suffering from a psychotic episode at the time of the offence absolving the individual of responsibility for the crime). Diminished responsibility, on the other hand, involves partial mitigation of perceived responsibility for criminal acts on the basis that a psychological disorder reduced the agent's ability to act in accordance with the law (e.g., a defendant suffering from posttraumatic stress disorder has their sentence downgraded from murder to manslaughter because it is believed the disorder leads to impaired judgement and/or impulse control). Hence, legal theory suggests that applying a psychiatric label will shift expectations about the sorts of social behaviours an agent is likely to exhibit and their perceived responsibility for certain negative actions. Applying this logic to the SEE paradigm, if a person acts in a way that leads to a negative social side-effect, attributions of intentionality may be reduced if that person was believed to have a psychiatric condition.

Psychological Models of Blame

Numerous psychological models of blame also take into account an agent's psychological condition when assessing culpability for immoral/illegal acts and omissions (e.g., Alicke, 2000; Heider, 1958; Weiner, 1995). For example, the culpable control model makes predictions about the factors which influence assessments of culpability (Alicke, 2000). However, this model also makes provisions for capacity constraints when assessing an agent's responsibility for bringing about an outcome (Alicke, 2000, p. 560; see also Gold, 2011, p. 527 for a discussion). Capacity constraints are factors that prevent an agent from acting in a desirable manner and/or diminish an agent's ability to refrain from undesirable behaviours and include psychological factors such as cognitive deficits (e.g., a poor understanding of the relevant social/behavioural norms) and emotional/psychiatric problems (e.g., psychological disorders such as depression and anxiety). Hence, these models provide a further rationale for examining the potential interaction between psychiatric status and the SEE.

Theoretical Accounts of the Side-Effect Effect

The present study used the "rational scientist" model proposed by Uttich and Lombrozo (2010) as a means of testing whether the SEE will be obtained for moral agents with a psychiatric disorder. According to the rational scientist model, information about whether side-effects conform to or violate social norms is the key driver of intentionality judgments (see also Papadopoulos and Hayes, 2018). Norm violating behaviours (e.g., wearing a suit to a casual picnic with friends), compared to norm conforming behaviours (e.g., wearing a suit to a job interview), license people to make stronger mental state and trait inferences about a moral agent. When the CEO helps the environment this behaviour is perceived as unintentional because it is attributed to the chairman following a widely held social norm to behave in a pro-environmental fashion whereas harming the environment violates this social norm and as such is perceived as intentional. Given that this model regards social normative expectations rather than moral considerations as key to the SEE, it predicts that moral agents with a psychiatric disorder will be assessed against a different set of social normative expectations. These moral agents will therefore be less likely to be judged as bringing about negative side-effects intentionally if the actions leading to this are consistent with what would be normatively expected of a person with a psychiatric disorder.

Other than the rational scientist model, intuitive moralist accounts (e.g., Knobe, 2003a, 2003b, 2006; see Feltz, 2007; Knobe, 2010 for reviews) are the main theoretical orientations to explaining the SEE. According to intuitive moralist accounts the key difference between harm and help SEE scenarios is the moral valence of the side-effects. Therefore, intentionality attributions for moral agents with a psychiatric disorder should remain largely unchanged according to intuitive moralist accounts because moral considerations relevant to the outcome (side-effect) are the key driver rather than an agent's psychiatric status.

The Present Studies

As in previous studies, agents brought about either positively or negatively valenced side-effects. However, the present studies used a novel paradigm inspired by the rational scientist model (Uttich and Lombrozo, 2010). Specifically, agents were given different psychiatric diagnostic labels as a means of manipulating norms about individual responsibility for social behaviours. By indicating that agents suffered from certain psychiatric disorders (or by withholding this information), side-effects could be manipulated as either conforming to or violating norms about responsibility for positive or negative social acts. This allowed for a novel test of whether the SEE would obtain for moral agents with a psychiatric disorder — the primary aim of the current series of studies. A secondary aim was to provide a test of the predictions of the rational scientist model and intuitive moralist accounts. In Experiment 1a moral agents were described as having a psychiatric disorder. Experiment 1b described participants as having an organic medical disorder that affected their behaviour. In Experiment 1c the psychiatric norms were explicitly stipulated. Experiment 2 replicated the first study but rather than introductory psychology students it tested participants with relative expertise in psychiatric norms.

Experiment 1a

This study employed novel written vignettes structurally similar to those used in prior studies of the SEE (e.g., Knobe, 2003a). The vignettes outlined a scenario where an agent acts to bring about some primary goal (e.g., to avoid attending a party to which he was invited). That primary goal gives rise to a foreseen but unintended side-effect (e.g., the agent's girlfriend misses out on going to the party which upsets her). Note that although the effect of the agent's behaviour impacts another person, this is still considered a side-effect because it is a foreseen but unintended consequence of the agent attempting to bring about their primary goal. As in previous studies, the moral valence of the side-effect was manipulated such that half the scenarios employed positively valanced side-effects (referred to as the help condition) and half employed negatively valanced side-effects (referred to as the harm condition).

A novel feature of this study was that the normative context in which the agent operated was also manipulated through the omission or addition of diagnostic information and a psychiatric diagnosis designed to alter whether the side-effects were perceived as norm conforming or norm violating. This meant that half the vignettes included diagnostic information and a psychiatric diagnosis for the agent (referred to as the *label-present* condition) and half did not include this diagnostic information (referred to as the *label-absent* condition).

The rational scientist model predicts that perceived intentionality for negatively valenced side-effects is mediated by the degree to which side-effects conform to or violate perceived social norms. If an agent acts in a manner consistent with social norms characteristic of some psychiatric disorder (e.g., an agent suffering from depression acts out of a sense of hopelessness and despondency) then the side-effect would be regarded as being consistent with the relevant social and behavioural norms (i.e., norm conforming). Hence, perceived intentionality for such side-effects (even when negatively valenced) should be relatively low.

Intuitive moralist accounts on the other hand claim that perceived intentionality is primarily a function of the moral content of side-effects. This means that the perceived norms relevant to an agent should have little bearing upon intentionality judgements. If an agent with a psychiatric disorder causes another person to suffer as a side-effect of their behaviour, intuitive moralist accounts would predict that perceived intentionality for negatively valenced side-effects would remain high. Note that in this scenario the impact upon the other person brought about by the agent is considered a side-effect because it is a foreseen but unintended consequence of the agent's behaviour aimed at bringing about their primary goal.

Therefore the predictions made by each model are: (1) the rational scientist model predicts a main effect of side-effect valence (i.e., an SEE), and a side-effect valence by psychiatric labelling interaction i.e., a robust SEE in the label-absent condition but an attenuation or elimination of the SEE in the label-present condition; (2) intuitive moralist accounts predict a significant main effect of side-effect valence (i.e., an SEE) but neither a main effect of psychiatric labelling nor an interaction between side-effect valence and psychiatric labelling is predicted.

Method

Participants

Participants were 64 introductory psychology students randomly selected from the student population (42 females) who participated for course credit. Ages ranged from 18 to 48 years (M = 19.73 years, SD = 4.06).

Design

This study employed a 2 (side-effect valence: help, harm) X 2 (psychiatric label: label-present, label-absent) between-subjects factorial design. Equal numbers of participants were randomly allocated to one of the four experimental conditions.

Materials

Four original vignettes were created (collectively referred to as the clinical vignettes). The structure of each vignette was modelled on those of Knobe (e.g., Knobe 2003a) and Uttich and Lombrozo (2010). Each vignette focused on a moral agent (the protagonist) and the effect of their behaviour on another character. The agents in each scenario suffered from a psychiatric disorder and acted out in a manner characteristic of that disorder. Four vignettes were constructed based on behaviours characteristic of social anxiety disorder, major depressive disorder, eating disorder, and bipolar disorder (manic episodes) as described in the DSM – 5 (American Psychiatric Association, 2013). The structure of each vignette was validated by consulting two practicing clinical psychologists and two trainee psychologists who reviewed the vignettes with a focus on their fit to the corresponding DSM diagnostic categories with consequent agreement across the board that each vignette provided an accurate and valid description of the target psychiatric diagnosis.

In each vignette the agent acted in order to achieve a primary goal, which also resulted in a side-effect. This side-effect had emotional consequences for another character in the scenario, either positive or negative. See Figure 1 for an example vignette. In one scenario, Arthur (the agent) has been invited to a party and his primary goal is to avoid going. The consequence of avoiding the party is that Arthur's girlfriend will miss out on seeing an old school acquaintance. This side-effect had either negative (harm version) or positive (help version) social and emotional consequences. For example, in the harm version the girlfriend misses out on seeing an old school friend which upsets her. In the help version the girlfriend avoids seeing an old school rival which pleases her.

Arthur has been invited to a birthday party. He would rather not attend but feels pressured to do so. Arthur knows that there will be many people at the party who he does not know and that concerns him. [Arthur worries that he will say something in a social situation that will result in his humiliation and he finds the possibility of being judged in a social situation distressing. Arthur was recently diagnosed with an anxiety disorder.]

Arthur's brother, David, approaches Arthur and tells him that, "I have a way of getting you out of having to go to the birthday party. I can provide the party-host with an excuse about you having to look after Mum. This will allow you to avoid going to the party, but it will also mean that Susanne (Arthur's girlfriend; who will only go to the party if Arthur does) will miss out on seeing an old school friend who she likes and who will be at the party, which will upset Susanne." {This will allow you to avoid going to the party, and it will also mean that Susanne (Arthur's girlfriend; who will only go to the party if Arthur does) will miss out on seeing an old school rival who she dislikes and who will be at the party, which will please Susanne."}

Arthur answered, "I do not care one way or another whether Susanne goes to the party and sees her old school friend. {Arthur answered, "I do not care one way or another whether Susanne goes to the party and sees her old school rival.} I just want to avoid going to the party myself. Go ahead and provide the party-host with an excuse."

David provided the party-host with an excuse which allowed Arthur to avoid going to the party. Sure enough, Susanne stayed at home with Arthur missing out on the opportunity to see her old school friend which upset Susanne. {Sure enough, Susanne stayed at home with Arthur missing out on the opportunity to see her old school rival which pleased Susanne.}

Figure 1: This vignette depicts an agent with an anxiety disorder (social anxiety disorder) and corresponds to the harm condition. The label-absent and label-present conditions are identical except the latter also contains the text shown at the end of the first paragraph (in square brackets). The text shown in the second, third, and last paragraphs (in curly brackets) corresponds to the analogous vignette in the help condition.

For those in the label-present condition the scenario was presented with an explicit diagnostic label (e.g., "Arthur was recently diagnosed with an anxiety disorder") and additional behavioural information relating to the psychiatric status of the agent (e.g., "Arthur worries that he will say something in a social situation that will result in his humiliation and he finds the possibility of being judged in a social situation distressing"). Those in the label-absent condition were not shown the label or additional psychiatric symptoms.

This study also employed two vignettes from previous studies of the SEE. These vignettes were included in order to establish that the SEE could be replicated with the sample of undergraduate students tested in this study. Specifically, the CEO vignette designed by Knobe (Knobe, 2003a) and the doctor vignette employed by Uttich and Lombrozo (2010, Experiment 3) were used (collectively referred to as the replication vignettes).

Procedure

Informed consent was first obtained from participants. Participants completed four clinical vignettes corresponding to one of the four experimental conditions (i.e., help/label-absent, help/label-present, harm/label-absent, harm/label-present). Each clinical vignette had an identical underlying structure. Participants were told they would be given a series of scenarios outlining different sorts of situations; they were asked some questions about each scenario and to provide the responses that feel most right for them.

The first four vignettes presented were always the clinical vignettes. The order in which these vignettes were presented was randomised across participants. For each vignette participants were asked four questions in a fixed order. These were adapted from Uttich and Lombrozo (2010, Experiment 2). Table 1 shows the four questions asked about the vignette corresponding to an agent suffering from an anxiety disorder, in the harm condition. Question 1 was the critical intentionality question (rating scale). Question 2 was a check to determine that we had successfully manipulated the perceived valence of the side-effects in the harm and help conditions. Question 3 had a similar aim and ascertained the extent to which an agent was perceived as being good or bad. Question 4 examined the extent to which an agent was perceived as being praiseworthy or blameworthy. These last two questions provided information about the impact of the manipulation of side-effect valence and the presence or absence of psychiatric labels on the perception of moral agents.

Table 1

Examples of Test Questions

Question	Purpose	Response Options
1: How appropriate would it be to say that Arthur intentionally upset Susanne?	Intentionality rating	1–7 scale: 1 = not at all appropriate, 4 = neither appropriate nor inappropriate, 7 = very appropriate
2: To what extent do you think the fact that Susanne missed out on going to the party and seeing her old school friend, which upset her, was good or bad?	Manipulation check of side-effect badness/goodness	1–7 scale: 1= bad, 4 = neither good nor bad, 7 = good
3: To what extent do you think that Arthur is a good or a bad person?	Moral agent badness/goodness	1–7 scale: 1= bad, 4 = neither good nor bad, 7 = good
4: To what extent do you think that Arthur should be blamed or praised?	Moral agent blame/praise	1–7 scale: 1= blamed , 4 = neither blamed nor praised, 7 = praised

After completing the four clinical vignettes participants were presented with the two replication vignettes. There were two versions of each replication vignette corresponding to the side-effect valence factor (i.e., help or harm) and participants were allocated to the same valence condition for the replication vignettes as they were for the clinical vignettes. The four questions asked after each of these vignettes followed the same structure as in the clinical vignettes.

Results

Clinical Vignettes: Preliminary Analyses of Intentionality Judgements with Each Vignette Analysed Separately

Intentionality ratings for the four clinical vignettes are shown in Table 2. These data are presented as a function of the four experimental conditions. The data met the requirements of parametric analyses including normality (data in each group were normally distributed), homoscedasticity (homogeneity of variance across groups), independence (data in each group were randomly and independently sampled from the population) and no outliers present. Preliminary analyses were conducted examining each clinical vignette separately in order to determine if each showed the predicted SEE. A one way ANOVA examining the effect of the valence (harm/help) manipulation was conducted separately for each vignette. Intentionality ratings were significantly greater in the harm than the help condition, averaged over the psychiatric label factor, for all four clinical vignettes; anxiety, F (1, 62) = 10.45, $p < .01$, $\eta_p^2 = .15$;[1] depression, F (1, 62) = 43.40, $p < .001$, $\eta_p^2 = .41$; eating disorder, F (1, 62) = 11.54, $p < .01$, $\eta_p^2 = .16$ and manic disorder, F (1, 62) = 23.91, $p < .001$, $\eta_p^2 = .28$. Hence, an SEE was obtained for each vignette on the rating measure. Given the consistent finding of a SEE across each of the four clinical vignettes, all subsequent analyses were collapsed across vignettes.

Table 2

Mean Ratings for the Intentionality Question in Each Experimental Condition[2]

	Experimental Condition			
Disorder	Help/No Label	Harm/No Label	Help/Label	Harm/Label
Anxiety	2.38 (1.82)	3.38 (1.09)	2.50 (1.41)	3.75 (1.18)
Depression	1.81 (1.22)	4.13 (1.36)	2.38 (1.20)	4.00 (0.97)
Eating Disorder	2.00 (1.41)	3.50 (1.32)	2.69 (1.49)	3.56 (1.36)
Manic Disorder	2.19 (1.60)	4.81 (1.68)	2.75 (1.81)	4.13 (1.41)
Aggregate	2.09 (1.34)	3.95 (0.75)	2.58 (1.04)	3.86 (0.73)

[1] This and all other effect sizes reported in this paper represent partial eta-squared values.

[2] For example, "How appropriate would it be to say that Arthur intentionally upset Susanne (1-7 scale; 1 = not at all appropriate, 7 = very appropriate)?" Standard deviations are shown in parentheses.

Clinical Vignettes: Intentionality Judgements (Aggregate Data for the Clinical Vignettes)

Mean intentionality ratings averaged across the four clinical vignettes are shown in Figure 2. These data were entered into a 2 (label) X 2 (valence) between-subjects ANOVA. The main effect of side-effect valence was significant, F (1, 60) = 39.72, $p < .001$, $\eta_p^2 = .40$. Intentionality ratings were significantly higher in the harm than the help condition, consistent with the SEE. As well as establishing an SEE, this study sought to determine if the presence of psychiatric information and labels attenuated the SEE. However, neither the main effect of the label factor nor the interaction between the side-effect valence and label factors were significant, both *Fs* < 1.5. Contrary to the predictions based on the rational scientist model, manipulation of psychiatric labels did not moderate the SEE.

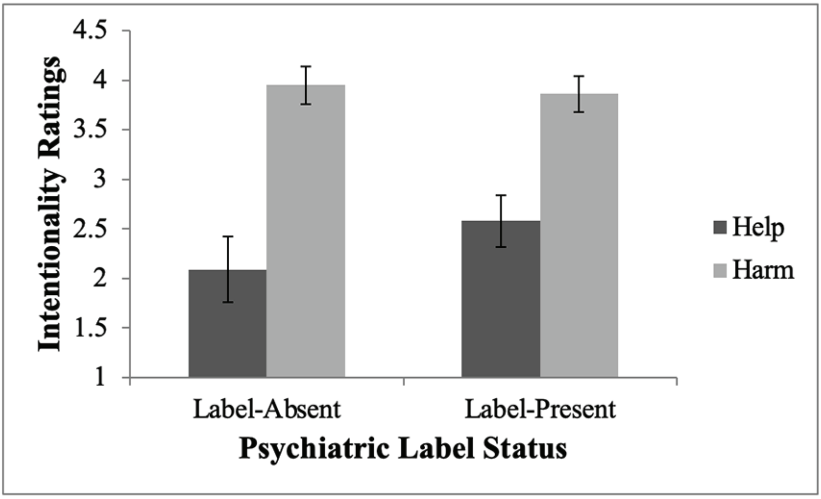

Figure 2: Mean intentionality ratings (with standard error bars) for the aggregated clinical vignettes as a function of psychiatric labels and side-effect valence.

Clinical Vignettes: Valence Check Questions (Aggregate Data for the Clinical Vignettes)

Mean ratings (averaged across vignettes) for the valence check questions are shown in Table 3. These data provide a key manipulation check of whether experimenter-defined help and harm side-effects were consistent with participants' valence perceptions. Each valence check question was analysed using a 2 (label) X 2 (valence) between-subjects ANOVA. For each valence check question: (1) participants rated side-effects as being worse in the harm than the help condition, F (1, 60) = 91.45, $p < .001$, $\eta_p^2 = .60$, however, neither the psychiatric label factor nor the valence by label interaction were significant, both *Fs* < 1.0; (2) participant perceptions of the degree to which agents were good or bad were unaffected by side-effect valence or a moral agent's psychiatric status, all *Fs* < 1.5; and (3) agents were rated as more blameworthy in the harm than in the help condition, *F* (1, 60) = 35.22, $p < .001$, $\eta_p^2 = .37$, however, there were no main effects or interaction involving the psychiatric label factor, both *Fs* < 1.0. Overall, the results for the valence check questions were generally consistent with expectations and suggest that the clinical vignettes provided a valid test of the SEE.

Table 3

Mean Ratings for the Three Valence Check Questions for the Aggregated Clinical Vignettes in Each Experimental Condition

Questions	Help/No Label	Harm/No Label	Help/Label	Harm/Label
Side-effect: Bad/Good	4.97 (1.05)	2.98 (1.01)	5.05 (0.80)	2.77 (0.65)
Moral Agent: Bad/Good	3.58 (0.85)	3.67 (0.85)	3.89 (0.85)	3.53 (0.49)
Moral Agent: Blame/Praise	3.92 (0.43)	3.16 (0.54)	3.83 (0.58)	3.06 (0.50)

Note: Each question used a seven-point response scale anchored as follows; side-effect: bad/good (1 = bad; 7 = good); moral agent: bad/good (1 = bad; 7 = good); moral agent: blame/praise (1 = blameworthy; 7 = praiseworthy). Standard deviations are shown in parentheses.

Replication Vignettes: CEO and Doctor Vignettes

The CEO and doctor vignettes were analysed separately. Mean ratings for the intentionality ratings are shown in Figure 3. One way analyses of variance found that intentionality ratings were significantly higher in the harm than the help condition for each vignette; CEO, *F* (1, 62) = 88.81, $\eta_p^2 = .59$; doctor, *F* (1, 62) = 17.14, $\eta_p^2 = .22$, both *ps* < .001.

Mean ratings for the three valence check questions for the CEO and doctor vignettes are shown in Table 4. The side-effect valence check questions confirmed that side-effects were rated as morally worse in the harm than in the help condition in both vignettes; CEO, *F* (1, 62) = 481.43, $\eta_p^2 = .89$; doctor, *F* (1, 62) = 161.79, $\eta_p^2 = .72$, both *ps* < .001. Ratings of the badness/goodness of the agent did not differ significantly across the harm and help conditions; CEO, *F* < 1.0; doctor, *F* < 1.5. However, participants were more likely to rate the agent as being blameworthy in the harm than in the help condition for both vignettes; CEO, *F* (1, 62) = 53.93, $p < .001$, $\eta_p^2 = .47$; doctor, *F* (1, 62) = 10.83, $p < .01$, $\eta_p^2 = .15$.

Taken together, the results for the valence check questions were largely consistent with expectations.

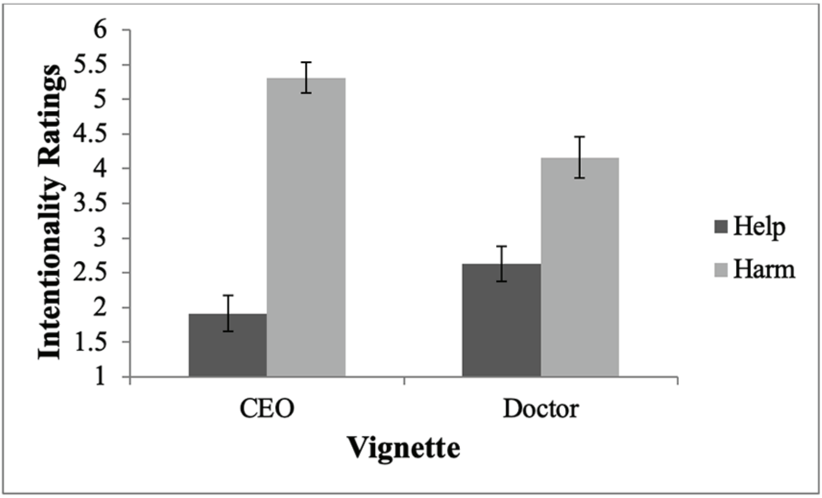

Figure 3: Mean intentionality ratings (with standard error bars) for the CEO and doctor vignettes as a function of side-effect valence; e.g., "How appropriate is it to say that the CEO intentionally harmed the environment (1 = not at all appropriate; 7 = very appropriate)?"

Table 4

Mean Ratings for the Valence Check Questions for the CEO and Doctor Vignettes as a Function of Side-Effect Valence

		Experimental Condition	
Vignette	Question	Help	Harm
CEO	Side-effect: Bad/Good	6.28 (1.08)	1.44 (0.62)
	Moral Agent: Bad/Good	2.66 (1.12)	2.38 (1.18)
	Moral Agent: Blame/Praise	3.44 (0.98)	1.72 (0.89)
Doctor	Side-effect: Bad/Good	6.06 (1.22)	2.38 (1.10)
	Moral Agent: Bad/Good	4.25 (1.55)	3.75 (1.85)
	Moral Agent: Blame/Praise	4.41 (1.32)	3.22 (1.56)

Note: Each question used a seven-point response scale anchored as follows: side-effect: bad/good (1 = bad; 7 = good); moral agent: bad/good (1 = bad; 7 = good); moral agent: blame/praise (1 = blameworthy; 7 = praiseworthy).

Discussion

The clinical vignettes yielded a significant SEE for intentionality rating data. However, neither the psychiatric label status main effect, nor the side-effect valence by psychiatric label status interaction, affected intentionality ratings. Hence, the SEE obtained across both the label-absent and label-present conditions. The SEE was also replicated using the CEO (Knobe, 2003a) and doctor (Uttich and Lombrozo, 2010) vignettes employed by prior studies.

The evidence suggests that people are insensitive to the normative information embedded in the clinical vignettes and appear to respond in a manner more consistent with intuitive moralist accounts (e.g., Knobe 2003a, 2006, 2010) than with the rational scientist model (Uttich and Lombrozo, 2010). The intuitive moralist account suggests that perceived intentionality is primarily driven by the moral content of side-effects and that consequently, perceived norms relevant to an agent should have little bearing upon intentionality judgements. However, such a conclusion may be premature on at least three grounds.

First, it is possible that the particular sample employed (i.e., introductory psychology students with little formal training in psychopathology) did not have an appropriate understanding of the characteristics of the psychiatric disorders used for this study. This would mean that participants did not have an adequate understanding of the relevant social and behavioural norms attached to the diagnostic labels.

Second, it could be the case that the link between psychiatric disorders and the specific behaviour of the moral agents in each vignette was not sufficiently salient. This would mean that even if participants had relevant norms about psychiatric disorders, that knowledge may not have been factored into intentionality judgements.

A third and related point is the possibility that psychiatric disorders may not be seen as a sufficiently compelling basis upon which to alter perceived intentionality in specific situations. This may be because the general social and behavioural norms associated with psychiatric disorders are weaker than the social norms relevant to the situations described in the vignettes (e.g., that it is rude for a guest not to eat dessert offered to her by the dinner party host). The subsequent studies attempt to address these issues. Experiment 1b sought to address the second and third issues and Experiment 1c the first issue outlined above.

Experiment 1b

Experiment 1b attempted to strengthen the influence of the labelling manipulation in two ways. First, participants were provided with information about a physical abnormality that the agent was suffering from as well as psychiatric diagnostic information about the agent. The rationale was that a medical condition may be perceived as a more compelling basis upon which to alter norms about social behaviour than a psychiatric condition. Some support for this argument comes from Ahn, Flanagan, Marsh, and Sanislow (2006) who found that undergraduate students with little formal training in psychology (like the participants in the present study) were more likely to perceive psychiatric conditions as having fewer defining features and causal essences than medical conditions. Further, Ahn et al. (2006) found that psychiatric conditions were more likely to be perceived as being social constructs created by relevant experts whereas (organic) medical conditions tended to be perceived as more "objectively" real. By specifying that the agent suffered from a physical abnormality (as well as a psychiatric condition) it was thought that participants may be more likely to change the social norms they apply when judging intentionality.

Second, the link between the agent's physical abnormality and their behaviour was made explicit. Highlighting the link between diagnostic status and behaviour provided another way of strengthening the test of the rational scientist model because it made the normative status of the behaviour relative to the diagnosis more salient. This labelling manipulation will be referred to as the *physical labelling condition*. The rational scientist model predicts an attenuation or elimination of the SEE in the physical labelling condition (compared to the label-absent condition in Experiment 1a) whereas intuitive moralist accounts predict that the SEE should be unaffected by the labels manipulation.

Method

Participants

Participants were 33 introductory psychology students (21 females) randomly selected from the student population who participated for course credit. Ages ranged from 17 to 23 years with a mean age of $M = 19.27$ years ($SD = 1.46$).

Design

This study employed a single factor with two levels (side-effect valence; help/physical label, harm/physical label) design. Participants were randomly allocated to the experimental conditions (help/physical label, $n = 17$; harm/physical label, $n = 16$). Intentionality ratings in the physical labelling conditions were compared to the label-absent condition in Experiment 1a.

Materials

The same materials as those employed by the help/label-present and harm/label-present conditions in Experiment 1a were used but with a few changes.

All these changes took place in the first paragraph of the vignettes. Unlike Experiment 1a, a physical abnormality was described as affecting the agent's behaviour. For example, participants were provided with the following additional information for the anxiety vignette: *"It has recently been discovered that Arthur has a non-lethal genetic abnormality that affects his behaviour. Arthur was recently diagnosed with an anxiety disorder."* See Figure 4 for an example vignette.

Procedure

The general procedure was similar to Experiment 1a. Participants were presented with the four clinical vignettes in random order and asked four questions for each vignette. However, participants were not presented with either of the replication vignettes.

> Arthur has been invited to a birthday party. He would rather not attend but feels pressured to do so. Arthur knows that there will be many people at the party who he does not know and that concerns him. *Arthur worries that he will say something in a social situation that will result in his humiliation and he finds the possibility of being judged in a social situation distressing. It has recently been discovered that Arthur has a non-lethal genetic abnormality that affects his behaviour. Arthur was recently diagnosed with an anxiety disorder.*

Figure 4: The first paragraph of the vignette describing a moral agent suffering from an anxiety disorder (social anxiety disorder). Those sections corresponding to the physical labelling condition are italicised.

Results and Discussion

Clinical Vignettes: Intentionality Judgements

Intentionality ratings in this study were collapsed across the four clinical vignettes and were compared to the aggregate data for the clinical vignettes in the label-absent control condition from Experiment 1a. These data are shown in Figure 5. A 2 (valence) X 2 (label condition) cross-experimental analysis was conducted on intentionality ratings. The main effect for the side-effect valence factor was significant, $F\ (1, 91) = 28.43$, $p < .001$, $\eta_p^2 = .24$. That is, intentionality ratings were significantly greater in the harm than in the help condition, signifying a SEE. However, the main effect of the labelling factor was not significant, $F\ (1, 91) < 1$. The interaction between the side-effect valence and labelling factors was not significant, $F\ (1, 91) = 2.70$, $p = .10$, $\eta_p^2 = .03$.

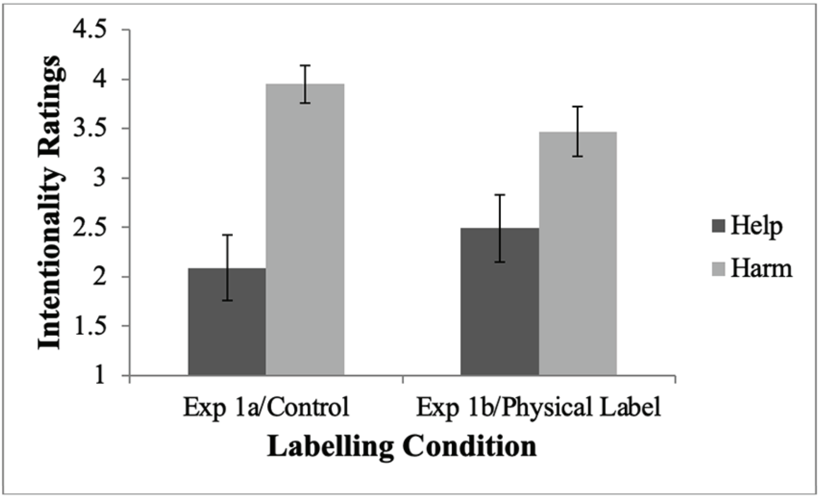

Figure 5: Mean intentionality ratings (with standard error bars) for the aggregated clinical vignettes as a function of side-effect valence for the control condition in Experiment 1a and the physical labelling condition in this study.

These results indicate that the SEE remained robust even when participants were presented with diagnostic information about an agent's physical illness and even when the link between the diagnostic information and an agent's behaviour was made explicit. Hence, perceived intentionality seemed to be unaffected by normative information relevant to an agent's psychiatric and physical diagnostic status. This challenges the prediction based on the rational scientist model. Given that perceived intentionality was sensitive to the moral valence of side-effects the evidence so far is tilted in favour of intuitive moralist accounts.

Experiment 1c

The lack of a labelling effect on perceived intentionality in the previous studies may have been due to participants having an inadequate understanding of the social and behavioural norms associated with the psychiatric diagnostic categories we employed. Experiment 1c attempted to further strengthen the labelling manipulation by explicitly stating the relevant norms. The approach of the present study is broadly similar to that of Uttich and Lombrozo (2010, Experiment 1) because both studies operationalized side-effect norm status by explicitly stipulating the norms. However, whereas the latter stipulated the norms numerically the present study did so by verbally stating the social/ behavioural norms relevant to the side-effects. This new labelling condition is referred to as the *explicit norms* *condition*. This manipulation should moderate the SEE if people take into account a psychiatric diagnosis when judging intentionality.

Method

Participants

Participants were 32 introductory psychology students (26 females) randomly selected from the student population who participated for course credit. Ages ranged from 17 to 26 years with a mean age of $M = 19.40$ years ($SD = 2.20$).

Design

This study employed a single factor with two levels (side-effect valence; help/explicit norms, harm/explicit norms) design. Equal numbers of participants were randomly allocated to one of the two experimental conditions. Results in the explicit norms condition were compared to the label-absent control from Experiment 1a.

Materials

The same materials as those employed by the label-present conditions in Experiment 1a were used, but with the following modifications to the first paragraph of each vignette. Notably, in this case the relevant behavioural norms for the psychiatric disorders were briefly summarised. Norms were always stipulated in such a way as to make it clear that someone with the psychiatric disorder in question could be expected to demonstrate different behaviours from someone without the disorder. Further, the specific stipulated norms were always directly relevant to the described side-effect. For example, for the anxiety vignette the following additional information was provided: *"Arthur has been diagnosed with an anxiety disorder. Scientific research has now established that people with an anxiety disorder are much more likely to worry about saying or doing something embarrassing in a social situation and so avoid social events more often than do people without an anxiety disorder."* See Figure 6 for an example vignette.

> Arthur has been invited to a birthday party. He would rather not attend but feels pressured to do so. Arthur knows that there will be many people at the party who he does not know and that concerns him. *Arthur has been diagnosed with an anxiety disorder. Scientific research has now established that people with an anxiety disorder are much more likely to worry about saying or doing something embarrassing in a social situation and so avoid social events more often than do people without an anxiety disorder.*

Figure 6: The first paragraph of the vignette describing an agent suffering from an anxiety disorder (social anxiety disorder). Those sections corresponding to the *explicit norms condition* are italicised.

Procedure

The general procedure was similar to Experiment 1b. Participants were presented with the four clinical vignettes in random order but were not presented with the replication vignettes.

Results and Discussion

Clinical Vignettes: Intentionality Judgements

Intentionality ratings for this study were collapsed across the four clinical vignettes and then compared to the aggregate data for clinical vignettes in the label-absent control condition from Experiment 1a. Intentionality ratings for the explicit norms and the label-absent control condition from Experiment 1a are shown in Figure 7. A 2 (valence) X 2 (label condition) cross-experimental analysis found a main effect of side-effect valence, F (1, 93) = 48.80, $p < .001$, $\eta_p^2 = .34$. Across label conditions, intentionality ratings were higher in the harm than in the help condition. Neither the main effect of labelling nor the label by valence interaction were significant, both $Fs < 1$. Therefore, a robust SEE was found for intentionality ratings, but the explicit statement of norms had little impact on perceived intentionality.

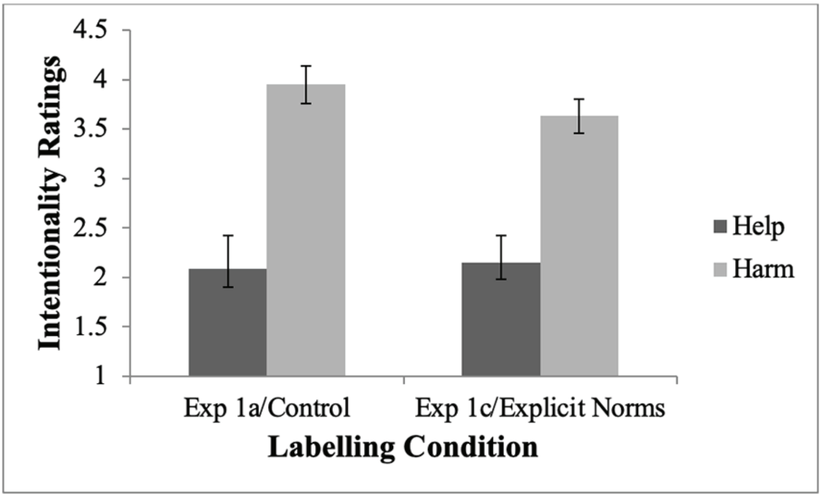

Figure 7: Mean intentionality ratings (with standard error bars) for the aggregated clinical vignettes as a function of side-effect valence for the control condition in Experiment 1a and the explicit norms condition in this study.

As in the previous two studies we found a reliable SEE but no effect of psychiatric labelling. This was the case even when social norms relevant to the psychiatric categories were explicitly stipulated. This suggests that participants were making their intentionality attributions based upon the moral valence of the side-effects rather than taking into account information about social and behavioural norms to inform their judgements. It is, however, possible that participants were factoring in information about norms relevant to psychiatric categories when judging intentionality, but that these norms were overshadowed by general norms relevant to the social situations outlined in the vignettes. Participants may have perceived general social norms about not upsetting other people (e.g., that it is wrong for a person to upset his girlfriend unnecessarily) as being more salient than those norms specific to the relevant psychiatric diagnosis. Hence, even if participants did take into account psychiatric norms when judging intentionality, these norms may not have been perceived as a sufficient basis to mitigate attributions of intentionality. Therefore, a further study was conducted to provide a stronger test of whether the SEE can be attenuated or eliminated for moral agents with a psychiatric disorder.

Experiment 2

Experiment 2 sought to address the possibility that participants were exhibiting an SEE despite taking into account information about norms relevant to psychiatric categories (due to the greater salience of general social norms) by using participants with formal training in psychopathology. Those with expertise in psychopathology should have a better understanding of the behavioural and social norms relevant to psychiatric diagnostic categories and therefore could be expected to place greater emphasis on these norms when making intentionality judgements.

This study replicated Experiment 1a with participants who had formal training in psychopathology and clinical experience. These participants were postgraduate students in the process of completing a postgraduate Master's program in clinical psychology.[3] Participants had experience in administering psychological treatments to clients in a university psychology clinic and had already completed at least one postgraduate course specifically focused on psychopathology including a detailed examination of psychiatric diagnostic categories, theories about aetiology, and psychological treatment approaches.

Examination of these "expert" participants should provide a stronger test of the labelling effects predicted by the rational scientist model. It was hypothesised

[3] This is the most common form of training for professional clinical psychologists in Australia. Doctoral level training is not required for professional clinical practice.

that for the clinical vignettes, (1) there would be a significant side-effect valence by labelling factor interaction; i.e., the SEE would be attenuated for experts in the *label-present condition* compared to experts in the *label-absent condition,* and (2) that experts would show a significant attenuation of the SEE in comparison to the novices tested in Experiment 1a. This latter prediction was tested using a cross-experimental comparison comparing data for the trainee clinicians from this study with those of the undergraduates from Experiment 1a. Participants in the current study were also administered the replication vignettes used in Experiment 1a and it was hypothesised that an SEE would obtain for these vignettes.

Method

Participants

Participants were 32 postgraduate students (29 females) completing either a higher degree research (n = 17) or coursework (n = 15) degree in clinical psychology (note that the clinical training component of these degrees is identical), randomly selected from the student population. Ages ranged from 22 to 45 years (M = 26.71 years, SD = 5.06).

Design and Procedure

This study employed the same materials, design and procedure as Experiment 1a. Participants were presented with the four clinical vignettes in random order, as well as both of the replication vignettes.

Results and Discussion

Clinical Vignettes: Intentionality Judgements

Mean intentionality ratings for the aggregate of the clinical vignettes are shown in Figure 8. A cross-experimental analysis was conducted comparing these data to analogous data collected from undergraduate “novices” in Experiment 1a. A 2 (side-effect valence: help, harm) X 2 (label condition: label-present, label-absent) X 2 (expertise: novices, experts) between-subjects ANOVA was carried out on intentionality ratings. There was a significant main effect of side-effect valence, F (1, 88) = 38.10, $p < .001$, $\eta_p^2 = .30$, consistent with the SEE. There was also a significant main effect of expertise, F (1, 88) = 16.79, $p < .001$, $\eta_p^2 = .16$. Those with postgraduate clinical training generally gave lower intentionality ratings than undergraduates. However, neither the main effect of labelling nor any of the two- or three-way interactions (including the interaction between expertise and side-effect valence) were significant, all *Fs* < 1. Hence, the SEE was not moderated by expertise level.

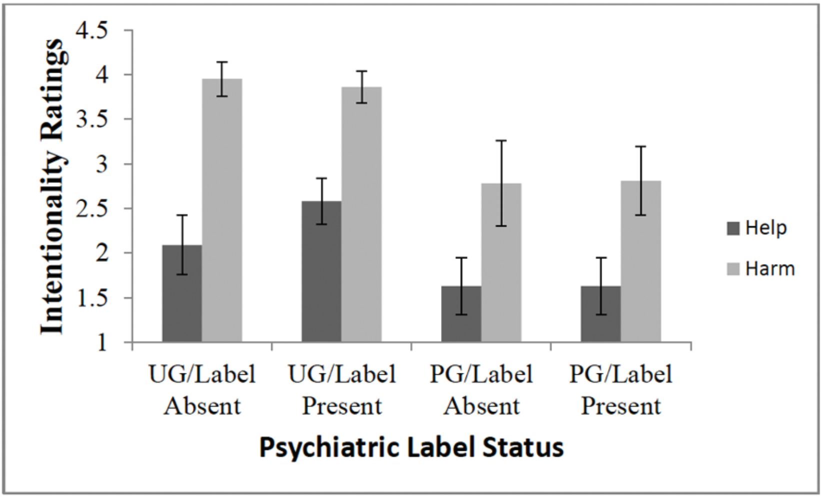

Figure 8: Mean intentionality ratings (with standard error bars) for the aggregated clinical vignettes as a function of side-effect valence and psychiatric labelling for undergraduates from Experiment 1a and postgraduates from this study (UG = undergraduates; PG = postgraduates).

Replication Vignettes: Intentionality Judgements

Responses to the CEO and doctor vignettes were analysed separately. Mean intentionality ratings are shown in Figure 9. A cross-experimental analysis was conducted to compare the responses to the replication vignettes in this study to those of Experiment 1a. A 2 (valence) X 2 (expertise) between-subjects ANOVA confirmed that for both scenarios there was a main effect of side-effect valence, consistent with the SEE: CEO, *F* (1, 30) = 102.95, $p < .001$, $\eta_p^2 = .53$; doctor, *F* (1, 30) = 16.88, $p < .001$, $\eta_p^2 = .16$. There was also a significant main effect of expertise for both scenarios: CEO, *F* (1, 30) = 7.41, $p < .01$, $\eta_p^2 = .08$; doctor, *F* (1, 30) = 4.12, $p < .05$, $\eta_p^2 = .04$. As was the case for the clinical vignettes, participants with postgraduate clinical training generally gave lower intentionality ratings than undergraduates. However, there was no interaction between side-effect valence and expertise for either scenario, both *Fs* < 1.

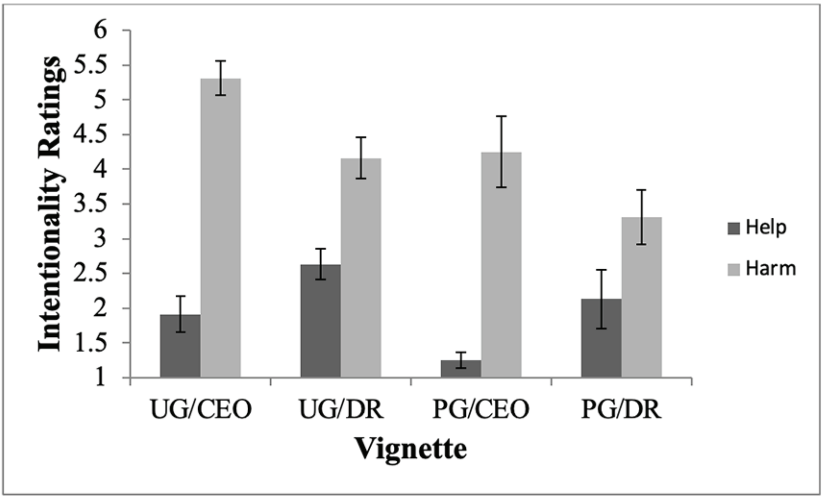

Figure 9: Mean intentionality ratings (with standard error bars) for the CEO and doctor vignettes as a function of side-effect valence for undergraduates from Experiment 1a and postgraduates from this study (UG = undergraduates; PG = postgraduates).

For the clinical vignettes a significant main effect of side-effect valence was found. However, neither the main effect of labelling nor the interaction between side-effect valence and labelling were significant. Hence, the SEE remained robust when psychiatric labels were applied to the agents. A comparison between this study and Experiment 1a (which tested undergraduate clinical "novices") found that experts gave lower intentionality ratings than novices although this effect was constant across all label and valence conditions. The SEE was also replicated with the CEO and doctor vignettes and once again experts gave lower intentionality ratings than novices across the board. Although experts were more conservative in attributing intentionality this was a general effect that applied regardless of the presence or absence of psychiatric labels. Indeed this general conservatism in attributing intentionality on the part of the postgraduate clinicians extended to the non-clinical vignettes.

General Discussion

The primary aim of the present series of studies was to examine whether the SEE would remain robust for judgments involving moral agents with a psychiatric disorder. A secondary aim was to provide a test of competing theoretical explanations of the SEE including the rational scientist model and intuitive moralist accounts. This series of studies used psychiatric diagnostic labels to manipulate social and behavioural norms to test whether the SEE is driven by normative rather than moral considerations. By manipulating whether or not moral agents were described as suffering from a psychiatric disorder these studies tested whether perceived intentionality for side-effects is sensitive to social normative information.

Experiment 1a tested both clinical vignettes and replication vignettes (i.e., CEO vignette, Knobe, 2003a; doctor vignette, Uttich and Lombrozo, 2010). The SEE was replicated with both sets of vignettes. However, there was no moderating influence of psychiatric diagnostic information on perceived intentionality. Similar results were found in Experiments 1b–1c. In these studies the SEE remained robust despite multiple attempts to strengthen the labelling manipulation. The use of more expert participants (Experiment 2) also failed to produce evidence of a moderation of the SEE when psychiatric labels were applied to agents although experts were generally more conservative in attributing intentionality despite showing a robust SEE. Taken together, the labelling studies did not support the prediction that labels which alter norms regarding individual responsibility for social actions alter the magnitude of the SEE. In this respect the data appear inconsistent with the rational scientist model. Nevertheless, this series of studies represents a robust demonstration of the SEE in a new domain using a novel paradigm.

Theoretical Implications

The rational scientist model proposes that people take into account social norms when judging intentionality and that moral considerations are only important in so far as they inform people about whether social norms have been violated, and consequently, will predict an attenuation or elimination of the SEE in the labelling conditions. Intuitive moralist accounts focus on the primacy of moral considerations and therefore predict no effect of the labelling manipulation. The present studies provided evidence consistent with intuitive moralist accounts. In these studies there was evidence of a robust SEE but no evidence of an attenuation of the effect when side-effect norm status was manipulated by applying psychiatric labels to agents. The insensitivity of intentionality judgements to the norms associated with psychiatric labels suggests that judgements were being determined by moral rather than normative considerations.

However, these findings cannot be regarded as definitive evidence against the rational scientist model. It is possible that participants were taking into account normative information but that moral norms overshadowed the psychiatric norms. Consistent with this possibility is evidence of improving knowledge of psychiatric disorders amongst the general public (Reavley and Jorm, 2012) and high rates of recognition of psychiatric symptomatology (depression) in tertiary students and staff (Reavley, McCann, and Jorm, 2012).

It is possible that the general social norms (e.g., "one should be good and avoid upsetting others where possible") were seen as a more compelling basis for assessing intentionality than the specific social/behavioural norms relevant to the psychiatric labels. Extensive evidence suggests that people treat social norms as a compelling basis for evaluating and reacting to others from a very young age (see Rakoczy and Schmidt, 2013 for a review). Turiel (1983) found that children from 3 to 4 years of age have already begun to understand the distinction between moral norms (i.e., norms based on considerations of justice and wellbeing) and conventional norms (i.e., norms that are socially constructed and consequently, somewhat arbitrary) and regard violations of moral norms as more serious. Moreover, children show a greater tendency to enforce moral rather than conventional norms on others (Schmidt, Rakoczy, and Tomasello, 2012). These findings demonstrate the early entrenchment of general social norms with moral content and may help to explain why the SEE persisted across each labelling manipulation and when testing experts.

Philosophy of Psychiatry

The findings of these studies are informative about issues relating to the philosophy of psychiatry. The fact that participants did not mitigate their intentionality attributions to negatively valenced actions performed by agents with a psychiatric diagnosis poses an interesting dilemma. On the one hand, individuals with a psychiatric disorder are often perceived as having diminished agency, and consequently, their ability to act as rational beings can be regarded as compromised. On the other hand, agents with a psychiatric disorder who perform harmful acts are still treated as being just as responsible for their actions as people without a psychiatric diagnosis (i.e., non-clinical agents). Hence, this represents a paradox between agents with a psychiatric diagnosis being treated as both *less capable* but *fully culpable* at the same time. This creates a number of risks including: (1) agents with a psychiatric diagnosis being treated unfairly, (2) cognitive dissonance in individuals who are asked to judge intentionality for actions performed by agents with a psychiatric disorder, and (3) the potential for individuals to act in bad faith when assessing the capability and culpability of agents with a psychiatric disorder. Hence, individuals may be willing to acknowledge the compromised agency of an agent with a psychiatric disorder so long as it does not interfere with other priorities such as the proclivity to apportion responsibility for blameworthy acts.

These risks must be considered in the context of the legal codes of many countries (for example, Western countries such as Australia, Canada, United Kingdom, United States, and Sweden) which stipulate that a defendant's psychiatric status should be taken into account when assessing guilt or innocence, or when considering sentencing for a defendant found guilty of a crime. Taken together, the findings of the present studies suggest a potential tension between the fact that although *in theory* a defendant's psychiatric status should be taken into account when assessing legal culpability and sentencing, *in practice* this may not always be the case. This is discussed further below.

Practical Implications

The present series of studies have important practical implications for how experts make intentionality judgments and jurors evaluate the culpability of defendants with a psychiatric diagnosis. The finding that experts were more conservative in attributing intentionality across both clinical and non-clinical scenarios was a general effect that applied whether or not psychiatric labels were present (Experiment 2). This conservatism in attributing intentionality may have been due to clinical trainees' greater awareness of the complex, multiple causal determinants of human behaviour. Consequently, these (relative) experts may have been less likely to attribute intentionality on the basis of the relatively limited information available in the vignettes. Indeed the evidence from the medical reasoning literature (e.g., Eva, 2002; Eva, Norman, Neville, Wood, and Brooks, 2002; Schmidt and Rikers, 2007) shows that medical trainees are more likely to use "analytical" reasoning when making clinical judgements which involves consciously sorting through and assessing the available information, hypothesis generation, and hypothesis testing. These findings suggest that trainee clinicians may demonstrate a similar "analytical" approach when making attributions about intentionality which may result in a more systematic search of the available information, a need for more information to support judgements and consequently, more conservative judgements.

Legal codes generally make allowances for jurors to take into account a defendant's mental state when assessing culpability for criminal acts. The SEE has been regarded as a potential problem for juror impartiality because it suggests moral considerations bias evaluations of culpability (e.g., Nadelhoffer, 2006) even when a defendant's psychiatric status should mitigate perceived culpability. The findings of the present study that perceived intentionality for negative side-effects continues to remain high even when moral agents have a psychiatric diagnosis further suggests that jurors may find it challenging to sufficiently mitigate assessments of criminal culpability for defendants with a psychiatric disorder despite the requirements of many legal codes. Future research can help to shed further light on this question by examining whether judgments made by mock jurors are sensitive to a defendant's mental state or psychiatric status by holding constant the nature of the crime whilst manipulating psychiatric status to determine if this effects a change in culpability judgments.

Limitations and Future Directions

This series of studies did not produce conclusive evidence of an attenuation of the SEE for moral agents with a psychiatric disorder. One major reason may have been the lack of calibration between the general social norms and the specific social or behavioural norms of the psychiatric diagnostic categories. Calibration between these competing norms refers to the equating of the perceived strength, salience, and importance of competing norms. Therefore, any future studies that wish to test whether the SEE can be attenuated or eliminated using this labelling paradigm will have to better calibrate these competing norms.

This could be done in two different ways. First, given that the psychiatric diagnoses were described as being of "moderate" severity, a labelling manipulation where agents are described as having more serious psychiatric conditions could be employed to increase the salience and strength of the social and cultural norms underpinning the disorder. For example, agents could be described as suffering from schizophrenia or autism spectrum disorder. Such labels may more effectively counteract the strength of competing general social norms as these conditions tend to involve symptoms or behaviours that represent a greater deviance from the non-clinical population than the disorders used in these studies (DSM – 5; American Psychiatric Association, 2013).

Second, participants could be primed to the relevance of psychiatric labels before making intentionality judgements. For example, a reminder of the relevance of the label could be given when asking the intentionality questions (e.g., "Given that Arthur has an anxiety disorder which makes him more likely to avoid social events, how appropriate is it to say that Arthur intentionally upset Susanne?").

Conclusion

The present studies examined whether the SEE obtains for moral agents with a psychiatric disorder and the validity of the rational scientist model as compared to intuitive moralist accounts employing a novel psychiatric labelling paradigm. The findings were consistent with intuitive moralist accounts. However, the possibility that an imbalance between the strength and salience of general social norms and specific psychiatric norms resulted in an insensitive test of the impact of psychiatric norms could not be ruled out. The present studies have extended our understanding of the SEE by providing a robust demonstration of the SEE in a new domain using a novel psychiatric labelling paradigm and demonstrating that expert clinical judgments of perceived intentionality are more generally conservative than those of laypeople.

References

Adams, F., and Steadman, A. (2004a). Intentional action in ordinary language: Core concept or pragmatic understanding? *Analysis, 64*, 173–181.

Adams, F., and Steadman, A. (2004b). Intentional actions and moral considerations: Still pragmatic. *Analysis, 64*, 268–276.

Ahn, W., Flanagan, E.H., Marsh, J.K., and Sanislow, C.A. (2006). Beliefs about essences and the reality of mental disorders. *Psychological Science, 17*, 759–766.

Alicke, M.D. (1992). Culpable control. *Journal of Personality and Social Psychology, 63*, 368–378.

Alicke, M.D. (2000). Culpable control and the psychology of blame. *Psychological Bulleti, 126*, 556–574.

Alicke, M.D. (2008). Blaming badly. *Journal of Cognition and Culture, 8*, 179–186.

American Psychiatric Association. (2013). *Diagnostic and satistical manual of mental disorders* (fifth edition). Arlington, Virginia: American Psychiatric Publishing.

Anckarsäter, H., Radovic, S., Svennerlind, C., Höglund, P., and Radovic, F. (2009). Mental disorder is a cause of crime: The cornerstone of forensic psychiatry. *International Journal of Law and Psychiatry, 32*, 342–347.

Buchanan, A., and Zonana, H. (2009). Mental disorder as the cause of a crime. *International Journal of Law and Psychiatry, 32*, 142–146.

Cokely, E.T., and Feltz, A. (2009). Individual differences, judgment biases, and theory-of-mind: Deconstructing the intentional action side effect asymmetry. *Journal of Research in Personality, 43*, 18–24.

Cushman, F., and Mele, A. (2008). Intentional action: Two-and-a-half folk concepts? In J. Knobe (Ed.), *Experimental philosophy* (pp. 171–188). Oxford: Oxford University Press.

Eva, K.W. (2002). The aging physician: Changes in cognitive processing and their impact on medical practice. *Academic Medicine, 77*, S1–S6.

Eva, K.W., Norman, G.R., Neville, A.J., Wood, T.J., and Brooks, L.R. (2002). Expert–novice differences in memory: A reformulation. *Teaching and Learning in Medicine: An International Journal, 14*, 257–263.

Feltz, A. (2007). The Knobe effect: A brief overview. *Journal of Mind and Behavior, 28*, 265–277.

Gold, A. (2011). Criminal culpability and self-control: Back to M'Naughton. *Psychiatry, Psychology and Law, 18*, 525–536.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Knobe, J. (2003a). Intentional action and side-effects in ordinary language. *Analysis, 63*, 190–193.

Knobe, J. (2003b). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology, 16*, 309–324.

Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies, 130*, 203–231.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences, 33*, 315–365.

Knobe, J., and Burra, A. (2006). The folk concept of intention and intentional action: A cross-cultural study. *Journal of Cognition and Culture, 6*, 113–132.

Leslie, A.M., Knobe, J., and Cohen, A. (2006). Acting intentionally and the side-effect effect: Theory of mind and moral judgment. *Psychological Science, 17*, 421–427.

Mitchell, E.W. (1999). Madness and meta-responsibility: The culpable causation of mental disorder and the insanity defence. *Journal of Forensic Psychiatry, 10*, 597–622.

Nadelhoffer, T. (2006). Bad acts, blameworthy agents, and intentional actions: Some problems for juror impartiality. *Philosophical Explorations, 9*, 203–219.

Nichols, S., and Ulatowski, J. (2007). Intuitions and individual differences: The Knobe effect revisited. *Mind and Language, 22*, 346–365.

Papadopoulos, C., and Hayes, B.K. (2018). What matters when judging intentionality — moral content or normative status? Testing the rational scientist model of the side-effect. *Psychonomic Bulletin and Review, 25*, 1170–1177.

Rakoczy, H., and Schmidt, M.F.H. (2013). The early ontogeny of social norms. *Child Development Perspectives, 7*, 17–21.

Reavley, N.J., and Jorm, A.F. (2012). Public recognition of mental disorders and beliefs about treatment: Changes in Australia over 16 years. *British Journal of Psychiatry, 200*, 419–425.

Reavley, N.J., McCann, T.V., and Jorm, A.F. (2012). Mental health literacy in higher education students. *Early Intervention in Psychiatry, 6*, 45–52.

Schmidt, M.F.H., Rakoczy, H., and Tomasello, M. (2012). Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition, 124*, 325–333.

Schmidt, H.G., and Rikers, R.M.J.P. (2007). How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Medical Education, 41*, 1133–1139.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, Massachusetts: Cambridge University Press.

Uttich, K., and Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition, 116*, 87–100.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford Press.