

A *stit* Logic Analysis of Morally Lucky and Legally Lucky Action Outcomes

Jan Broersen

Abstract Moral luck is the phenomenon that agents are not always held accountable for performance of a choice that under normal circumstances is likely to result in a state that is considered bad, but where due to some unexpected interaction the bad outcome does not obtain. We can also speak of ‘moral misfortune’ in the mirror situation where an agent chooses the good thing but the outcome is bad. This paper studies formalizations of moral and legal luck (and moral and legal misfortune). The three ingredients essential to modelling luck of these two different kinds are (1) indeterminacy of action effects, (2) determination on the part of the acting agent, (3) the possibility of evaluation of acts and/or their outcomes relative to a normative moral or legal code. The first, indeterminacy of action, is modelled by extending *stit* logic by allowing choices to have a probabilistic effect. The second, deliberateness of action, is modelled by (a) endowing *stit* operators with the possibility to specify a lower bound on the change of success, and (b) by introducing the notion of attempt as a maximisation of the probability of success. The third, evaluation relative to a moral or legal code, is modelled using Anderson’s reduction of normative truth to logical truth. The conclusion will be that the problems embodied by the phenomenon of moral luck may be introduced by confusing it with legal luck. Formalizations of both forms are given.

1 Introduction

Agents may be morally lucky (Williams 1982) in several ways. Nagel (1979) explains how an agent can be morally lucky due to circumstance: if circumstances would have been different, for instance, if an agent would have had the opportunity to steal

J. Broersen (✉)

Intelligent Systems Group, Department of Information and Computing Sciences,
Faculty of science, Universiteit Utrecht, P.O. Box 80089, 3508 TB UTRECHT, The Netherlands
e-mail: broerser@cs.uu.nl

somebody's money without anybody noticing, then the agent would have done so and would have gone morally wrong. Another category is moral luck due to character, such as when an agent would have committed an immoral act of retribution in case it would be less forgiving than it actually is. But these classes of moral luck are not as interesting, in my opinion, as the class related to non-determinate outcomes of actions.¹ That is, cases where an agent is morally lucky because his agency interferes with nature or the agency of other agents in such a way that its immoral behavior does not lead to an outcome that is considered morally bad.

If moral luck has to be taken serious as a principle of ethical reasoning, that is, if we really agree that, at least to a certain extent, lucky outcomes secure the acting agent from blame, then moral luck presents us with a problem. This problem is that moral luck conflicts with another principle of ethical reasoning, namely, that agents are not morally responsible for actions or outcomes that are not under their control. A case where this conflict between moral principles comes to the foreground, is the following variant on the well-known drinking and driving examples.

Example 1.1 *In scenario 1 a man drinks 10 beers. He drives home, knows he is taking a risk, and drives very carefully. After 1 km, he is held by the police. The alcohol percentage in his blood is measured to be 0.15%. He gets fined 300 Euros. Scenario 2 is only slightly different. A man drinks 10 beers. He drives home, knows he is taking a risk, and drives very carefully. However, now after 1 km, he fatally hits somebody crossing the road. After the accident the alcohol percentage in his blood is measured to be 0.15%. He gets convicted for involuntary manslaughter.*

If the moral responsibility in both scenarios of Example 1.1 is different, than this justifies the difference in legal treatment of the two scenarios; the view would be that driving drunk and killing somebody is morally despicable and is better than driving drunk without hitting anybody. However, the problem is that it is not clear that indeed moral responsibility is different in the two scenarios. The example is designed to make it clear that in scenario 1 and 2 both agents took the same *risk*; until 1 km of driving, the period in which the agents exercised their agency, the two scenarios are entirely identical. And, that the scenarios are indeed identical is to a certain extent also proven to a judge having to decide on the verdicts in both scenarios, since there is objective evidence that in both cases the percentage of alcohol was 0.15%. So if the agency involved in both scenarios is identical, and if moral responsibility is proportional to the agency or control exercised over a situation, the conclusion should be that the moral responsibility in both scenarios is the same. The conflict is then between two principles: if we commit ourselves to the standpoint that there is a difference in moral responsibility for the two scenarios, than this difference must be linked to the different outcomes of the scenarios. So, in that case we commit ourselves to the principle of (the possibility of) moral luck. The second, conflicting principle is that agents cannot be morally responsible for what is not under their

¹ I do appreciate however Nagel's argument that if we try to deal with the problem of moral luck relative to action effects by simply denying it, the only thing we do is to push the problem to deeper levels of consideration like those concerning character or intention.

control, which implies that, given their set-up, there should not be a difference in assignment of moral responsibility between the two scenarios.

In this article, in Sects. 4, 5 and 6, I will defend the standpoint that moral luck with respect to action outcomes, as described above, does not exist. I will then explain the difference in legal treatment of the two scenarios as resulting from differences between the legal and the moral evaluation of actions: in a moral evaluation they are identical, in a legal evaluation they are not. After all, legal responsibility is not the same as moral responsibility. A father can be legally responsible for wrongdoings of his children without being morally responsible for these same wrongdoings. Or a company can be legally responsible for polluting the environment without being morally responsible (because it is not clear at all if moral responsibility can be lifted to groups or organisations of agents). The problem with this 'way out' is that it raises the question of how much moral and legal responsibility can diverge. It is clear that legal responsibility always reflects at least some underlying level of moral responsibility. First of all, our laws cannot be too different from what people believe to be morally justified, otherwise our social choice mechanisms will correct them. Second, to a certain extent the father is morally responsible, since in a remote way, being responsible for their upbringing, his children's acts derive from his own acts. And the company is maybe also morally responsible in some derived sense since somehow the company's acts are acts of the agents working for the company. These observations ask for clarification, the kind of clarification that in my opinion can best be provided by formalisation.

Before explicating the goal of this paper further, I want to take away a possible misconception. The described problem with moral luck is mostly seen as conflicting with Kantian ethics. Kantian ethics links morality to the action and not to its outcome. So, if in practice we assign morality to outcomes instead of actions, as we seem to do in many cases, we are operating in conflict with Kantian ethics. It might seem that we have to conclude that our practice of assigning morality to outcomes is an argument in favour of consequentialist ethics. However, that would be too hasty a conclusion. Moral luck is just as much a problem for consequentialist ethics as it is for Kantian ethics. Consequentialist ethics takes the position that badness of acts *derives* from the badness of their outcomes. But this dependency on the status of outcomes does not transfer to cases where outcomes are uncertain, that is, to cases where moral luck can be involved. For instance, a consequentialist can argue that his killing is justified or even the good thing to do if it saves more lives than it costs. Now also for the ethical reasoning in judging this consequentialist killer moral luck poses a problem, since if it turns out that this agent's killing does *not* save the lives of several others, the moral luck (in this case moral misfortune) phenomenon will raise its head and the killer will have a hard time justifying his killing. That is, the killing agent is likely to be judged for the unlucky outcome even though before the killing he was correct to expect that his act would save more lives than it would cost and was therefore, according to his consequentialist ethical reasoning, the good thing to do.

I strongly believe that we can get a much clearer picture of the situation by giving formalizations of agency, failure, attempt, negligence and (normative) luck. So this

is what I aim to do in this paper. In order to formalize luck in a normative context, we should resolve three core issues. We need:

1. a way to represent indeterminacy of action
2. a way to express the determination of an agent (risk avoidance, risk taking, negligent acting or refraining, attempt, etc.)
3. a way to represent the moral or legal value of an act relative to some (implicit) normative code.

The first problem we solve by resorting to probabilistic *stit* (Broersen 2011c). In this form of *stit* theory, effects of choices are no longer guaranteed but are obtained with a certain probability. The second problem we solve by endowing the probabilistic *stit* operators with lower bounds on the chance of success and by defining attempt as a maximization of the chance of success (Broersen 2011a). Finally, the third problem is tackled by introducing violation constants in this context (Anderson 1958; Broersen 2011b). If a violation occurs, an agent does not behave according to some moral or legal code that, in this paper, is not made explicit. We can see the paper then as a combination of the ideas put forward in Broersen (2011a, b, c). We will present essential parts of the material from those papers here and then discuss the application to the formalization of moral and legal luck.

2 Modeling Indeterminacy of Action

In the following two sub-sections we first introduce the base *stit* logic and then extend this logic to allow for non-determinate effects.

2.1 Determinate Action: $XSTIT^P$

In this section we define the base logic, which is a variant of the logic $XSTIT$ that we call $XSTIT^P$ (Broersen 2013). The difference with $XSTIT$ is embodied by an axiom schema concerning modality-free propositions p , which explains the name. The semantics uses h -relative effectivity functions, which specialize the notion of effectivity function from Coalition Logic (Pauly 2002) by defining choices relative to histories.

Definition 2.1 *Given a countable set of propositions P and $p \in P$, and given a finite set Ags of agent names, and $ag \in Ags$, the formal language \mathcal{L}_{XSTIT^P} is:*

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid [ag \text{ xstit}]\varphi \mid X\varphi$$

Besides the usual propositional connectives, the syntax of $XSTIT^P$ comprises three modal operators. The operator $\Box\varphi$ expresses ‘historical necessity’, and plays

the same role as the well-known path quantifiers in logics such as CTL and CTL* (Emerson 1990). Another way of talking about this operator is to say that it expresses that φ is ‘settled’. We abbreviate $\neg\Box\neg\varphi$ by $\Diamond\varphi$. The operator $[ag \ xstit]\varphi$ stands for ‘agent ag sees to it that φ in the next state’. We abbreviate $\neg[ag \ xstit]\neg\varphi$ by $\langle ag \ xstit \rangle\varphi$. The third modality is the next operator $X\varphi$. It has a standard interpretation as the transition to a next state.

Definition 2.2 An *XSTIT^P*-frame is a tuple $\langle S, H, E \rangle$ such that²:

1. S is a non-empty set of static states. Elements of S are denoted s, s' , etc.
2. H is a non-empty set of ‘backwards bundled’ histories. A history $h \in H$ is a sequence $\dots s, s', s'' \dots$ of mutually different elements from S . To denote that s' succeeds s on h we use a successor function $succ$ and write $s' = succ(s, h)$. The following constraint on the set H ensures that if different histories share a state, they are bundled together in the past direction:
 - a. if $s = succ(s', h)$ and $s = succ(s'', h')$ then $s' = s''$
3. $E : S \times H \times Ags \mapsto 2^S$ is an h -effectivity function yielding for an agent ag the set of next static states allowed by the choice exercised by the agent relative to a history. We have the following constraints on h -effectivity functions:
 - a. if $s \notin h$ then $E(s, h, ag) = \emptyset$
 - b. if $s' \in E(s, h, ag)$ then $\exists h' : s' = succ(s, h')$
 - c. if $s' = succ(s, h')$ and $s' \in h$ then $s' \in E(s, h, ag)$
 - d. $E(s, h, ag_1) \cap E(s, h', ag_2) \neq \emptyset$ for $ag_1 \neq ag_2$.

In Definition 2.2 above, we refer to the states s as ‘static states’. This is to distinguish them from ‘dynamic states’, which are combinations $\langle s, h \rangle$ of static states and histories. Dynamic states function as the elementary units of evaluation of the logic. This means that the basic notion of ‘truth’ in the semantics of this logic is about dynamic conditions concerning choices. This distinguishes *stit* from logics like Dynamic Logic and Coalition Logic whose central notion of truth concerns static conditions holding for static states.

The name ‘ h -effectivity functions’ for the functions defined in item **3** above is short for ‘ h -relative effectivity functions’. This name is inspired by similar terminology in Coalition Logic whose semantics is in terms of ‘effectivity functions’. Condition **3.a** above states that h -effectivity is empty for history-state combinations that do not form a dynamic state. Condition **3.b** ensures that next state effectivity as seen from a current state s does not contain states s' that are not reachable from the current state through some history. Condition **3.c** expresses the well-known *stit* condition of ‘no choice between undivided histories’. Condition **3.d** above states that simultaneous choices of different agents never have an empty intersection. This represents a constraint on the independence of choices by different agents.

² In the meta-language we use the same symbols both as constant names and as variable names, and we assume universal quantification of unbound meta-variables.

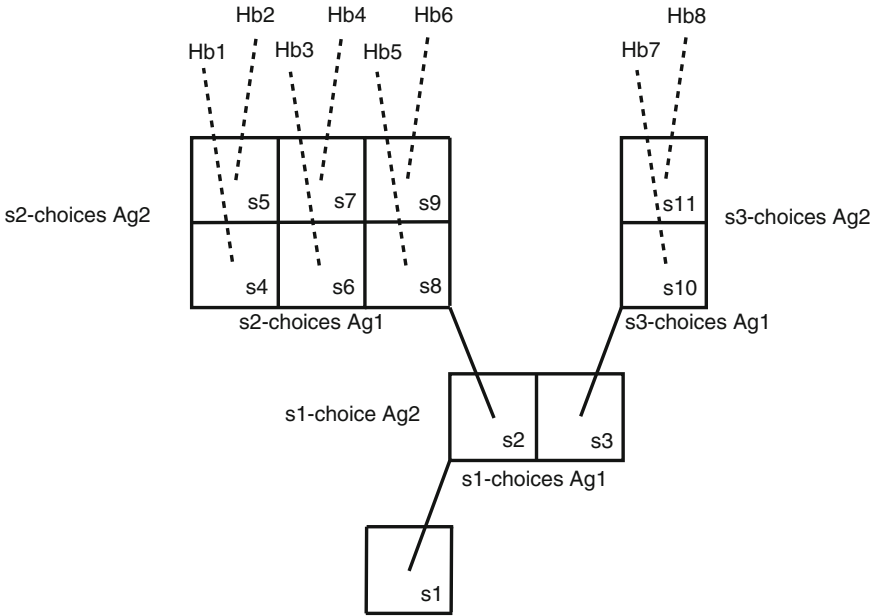


Fig. 1 Visualization of a partial two agent $XSTIT^P$ frame

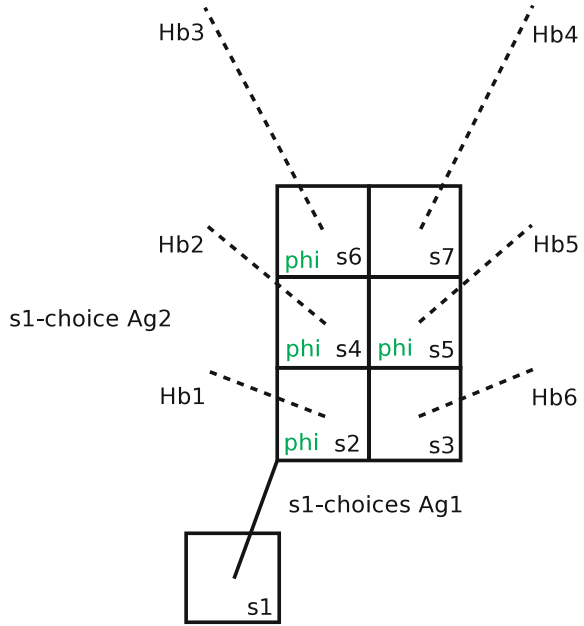
Figure 1 visualizes a frame of the type defined by Definition 2.2. The columns in the games forms linked to each state are the choices of agent ag_1 and the rows are the choices of agent ag_2 . Independence of choices is reflected by the fact that the game forms contain no ‘holes’ in them. Choice taking in this ‘bundled’ semantics is thought of as the separation of two bundles of histories: one bundle ensured by the choice exercised and one bundle excluded by that choice. The pictures of the frames suggest more constraints than are actually specified by Definition 2.2. For instance, the technical definition of the frames does not exclude that the choices of an agent ag are mutually disjoint. However, since they result in much tidier pictures, in the visualizations of the frames we assume such conditions.

We now define models by adding a valuation of propositional atoms to the frames of Definition 2.2. We impose that all dynamic states relative to a static state evaluate atomic propositions to the same value. This reflects the intuition that atoms, and modality-free formulas in general do not represent dynamic information. Their truth value should thus not depend on a history but only on the static state. This choice does however make the situation non-standard. It is a constraint on the models, and not on the frames.

Definition 2.3 A frame $\mathcal{F} = \langle S, H, E \rangle$ is extended to a model $\mathcal{M} = \langle S, H, E, V \rangle$ by adding a valuation V of atomic propositions:

- V is a valuation function $V : P \rightarrow 2^S$ assigning to each atomic proposition the set of static states relative to which they are true.

Fig. 2 Visualization of a partial two agent XSTIT^p model



We evaluate truth with respect to dynamic states built from a dimension of histories and a dimension of static states.

Definition 2.4 *Relative to a model $\mathcal{M} = \langle S, H, E, V \rangle$, truth $\langle s, h \rangle \models \varphi$ of a formula φ in a dynamic state $\langle s, h \rangle$, with $s \in h$, is defined as:*

$$\begin{aligned}
 \langle s, h \rangle \models p & \Leftrightarrow s \in V(p) \\
 \langle s, h \rangle \models \neg\varphi & \Leftrightarrow \text{not } \langle s, h \rangle \models \varphi \\
 \langle s, h \rangle \models \varphi \wedge \psi & \Leftrightarrow \langle s, h \rangle \models \varphi \text{ and } \langle s, h \rangle \models \psi \\
 \langle s, h \rangle \models \Box\varphi & \Leftrightarrow \forall h' : \text{if } s \in h' \text{ then } \langle s, h' \rangle \models \varphi \\
 \langle s, h \rangle \models X\varphi & \Leftrightarrow \text{if } s' = \text{succ}(s, h) \text{ then } \langle s', h \rangle \models \varphi \\
 \langle s, h \rangle \models [ag \text{ xstit}]\varphi & \Leftrightarrow \forall s', h' : \text{if } s' \in E(s, h, ag) \text{ and } \\
 & \quad s' \in h' \text{ then } \langle s', h' \rangle \models \varphi
 \end{aligned}$$

Satisfiability, validity on a frame and general validity are defined as usual.

Note that the historical necessity operator quantifies over one dimension, and the next operator over the other. The *stit* modality combines both dimensions. Now we proceed with the axiomatization of the base logic.

Figure 2 gives an example model that we can use to discuss the evaluation of formulas. Relative to static state s_1 and the history h_5 that is part of the bundle of histories Hb_5 we do not have that the choice by agent ag_1 ensures that φ holds, since the other agent has two choices (the bottom one and the top one) for which φ will not be true. So in this model we have that $\langle s_1, h_5 \rangle \not\models [ag_1 \text{ xstit}]\varphi$. However, relative

to, for instance, a history in the bundle Hb_1 , the agent ag_1 does guarantee that φ obtains as the result of the choice it exerts independent of what agent ag_2 choses simultaneously: for all three choices of the other agent φ is the result. So we have that $\langle s_1, h_1 \rangle \models [ag_1 \text{ xstit}]\varphi$.

Definition 2.5 *The following axiom schemas, in combination with a standard axiomatization for propositional logic, and the standard rules (like necessitation) for the normal modal operators, define a Hilbert system for $XSTIT^p$:*

- (p) $p \rightarrow \Box p$ for p modality free
S5 for \Box
- (D) $\neg[ag \text{ xstit}]\perp$
- (Lin) $\neg X\neg\varphi \leftrightarrow X\varphi$
- (Sett) $\Box X\varphi \rightarrow [ag \text{ xstit}]\varphi$
- (XSett) $[ag \text{ xstit}]\varphi \rightarrow X\Box\varphi$
- (Agg) $[ag \text{ xstit}]\varphi \wedge [ag \text{ xstit}]\psi \rightarrow [ag \text{ xstit}](\varphi \wedge \psi)$
- (Mon) $[ag \text{ xstit}](\varphi \wedge \psi) \rightarrow [ag \text{ xstit}]\varphi$
- (Dep) $\Diamond[ag_1 \text{ xstit}]\varphi \wedge \dots \wedge \Diamond[ag_n \text{ xstit}]\psi \rightarrow$
 $\Diamond([ag_1 \text{ xstit}]\varphi \wedge \dots \wedge [ag_n \text{ xstit}]\psi)$
for $Ags = \{ag_1, \dots, ag_n\}$

Theorem 2.1 (Broersen 2011b) *The Hilbert system of Definition 2.5 is complete with respect to the semantics of Definition 2.4.*

2.2 Action with Non-Determinate Effect: $XSTIT.Prob$

The *stit* logic of the previous section was based on the idea of acting as ensuring a certain condition. In the present section we put forward a theory that relaxes this assumption. Now actions are no longer necessarily successful. We are going to assume we measure success of action against an agent's beliefs about the outcome of its choice. So, the perspective is an internal, subjective one, and the criterion of success is formed by an agent's beliefs about its action. To represent these beliefs we choose here to use probabilities. In particular, we will represent beliefs about simultaneous choices of other agents in a system as subjective probabilities. Several choices have to be made. We will assume that an agent can never be mistaken about its own choice, but that it can be mistaken about choices of others. The actual action performed results from a simultaneous choice of all agents in the system. Then, if an agent can be mistaken about the choices of other agents (including possibly an agent with special properties called 'nature'), the action can be unsuccessful.

We introduce operators $[ag \text{ xstit}^c]\varphi$ with the intended meaning that agent ag exercises a choice for which it believes to have a chance of at least c of bringing about φ . We assume that numbers c are between 1 and 0 and that the set of possible c 's is at least countable (that is, a subset of \mathbb{Q}). Roughly, the semantics for this new operator

is as follows. We start with the multi-agent *stit*-setting of the previous section. Now to the semantic structures we add functions such that in the little game-forms, as visualized by Fig. 1, for each choice of an agent ag we have available the subjective probabilities applying to the simultaneous choices of other agents in the system. For an agent ag the sum of the probabilities over the choices of each particular other agent in the system adds up to one. So, the probabilities represent agent ag 's beliefs concerning what choices are exercised simultaneously by other agents. We use this subjective probability function to define for each choice the change of success to obtain a condition φ : relative to the choice we add up the probabilities for each of the choices of all other agents in the system leading to a situation obeying φ .

For the definition of the probabilistic frames, we first define an augmentation function returning the choices a group of agent has in a given state.

Definition 2.6 *The range function $Range : S \times Ags \mapsto 2^{2^S \setminus \emptyset} \setminus \emptyset$ yielding for a state s and an agent ag , the choices this agent has in s is defined as:*

$$Range(s, ag) = \{Ch \mid \exists h : s \in h \text{ and } Ch = E(s, h, ag)\}$$

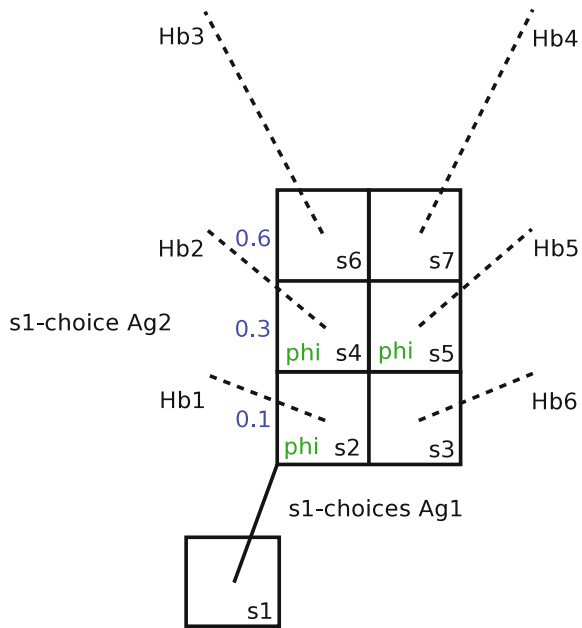
A range function is similar to what in Coalition Logic (Pauly 2002) is called an ‘effectivity function’. Now we are ready to define the probabilistic *stit* frames.

Definition 2.7 *A probabilistic $XSTIT^p$ -frame is a tuple $\langle S, H, E, B \rangle$ such that:*

1. $\langle S, H, E \rangle$ is an $XSTIT^p$ -frame
2. $B : S \times Ags \times Ags \times 2^S \mapsto [0, 1]$ is a subjective probability function such that $B(s, ag_1, ag_2, Ch)$ expresses agent 1's belief that in static state s agent 2 performs a choice resulting in one of the static states in Ch . We apply the following constraints.
 - a. $B(s, ag, ag', Ch) = 0$ if $ag \neq ag'$ and $Ch \notin Range(s, ag')$
 - b. $B(s, ag, ag', Ch) > 0$ if $ag \neq ag'$ and $Ch \in Range(s, ag')$
 - c. $\sum_{Ch \in Range(s, ag')} B(s, ag, ag', Ch) = 1$ if $ag \neq ag'$
 - d. $B(s, ag, ag, Ch) = 1$.

The conditions in Definition 2.7 are variations on the Kolmogorov axioms for probability. Condition 2.a says that agents only assign non-zero subjective probabilities to choices other agents objectively have. Condition 2.b says these probabilities are strictly larger than zero. Condition 2.c says that the sum of the subjective probabilities over the possible choices of other agents add up to 1. Condition 2.d says that agents always know what choice they exercise themselves. We may call this property the ‘free will’ property. In an *objective* view on the choices of an agent, the probabilities for the choices in the agent's repertoire have to add up to 1. From such an objective view point for each of the possible choices we get a chance somewhere between 0 and 1, and the standard Kolmogorov conditions apply. But from the perspective of the agent itself, that is, from the subjective viewpoint taken in this logic, the standard conditions do not apply. Conditional on what choice an agent actually takes, we can say that subjectively, the agent is 100% sure about what choice it is

Fig. 3 Visualization of a partial two agent probabilistic XSTIT^P model



taking. And at the same time it has the free will to take any of the choices open to him/her. No agent would regard its own choice taking as a matter of chance: it has the free will to choose and if it chooses something it is 100 % sure about the fact that it is taking that choice. That is what free will demands. So from a subjective viewpoint and taking into account the free will of the agents, we have reason to violate the third Kolmogorov axiom and allow for the possibility that an agent’s subjective probabilities concerning its own possibilities for taking choices add up to infinity (if the number of choices is infinite).³

Figure 3 extends the earlier example model with subjective probabilities for agent *ag1*’s belief concerning the choice agent *ag2* exercises simultaneously. We see that agent *ag1* believes the the chance that agent *ag2* chooses the top row is 0.6, that the chance for the middle row is 0.3 and the chance for the bottom row is 0.1. It is easy to check that this model satisfies all the conditions discussed above.

In the sequel we will need an augmentation function yielding for an agent and an arbitrary next static state the chance an agent ascribes to the occurrence of this state (given its belief, i.e., subjective probabilities about simultaneous choice taking of other agents). For this, we first need the following proposition. To guarantee that the proposition is true, we need the extra condition that choices do not overlap, which we can safely add to the semantics.

³ We can also give brief explanations of the determinist and compatibilist positions in this context: a determinist would argue that objectively, the chance for one of the choices is one while for the other choices, they are 0. A compatibilist would then argue that this is compatible with the agent assigning itself probability 1 to any of the choices: it cannot know that its choice is actually determined.

Proposition 2.2 *For any pair of static states s and s' for which there is an h such that $s' = \text{succ}(s, h)$ there is a unique ‘choice profile’ determining for each agent ag in the system a unique choice $Ch = E(s, h, ag)$ relative to s and s' .*

Now we can define the subjective probabilities agents assign to possible system outcomes. Because of the idea of independence of choices, we can multiply the chances for the choices of the individual agents relative to the system outcome (the resulting static state). Note that this gives a new and extra dimension to the notion of (in)dependence that is not available in standard *stit* theories.⁴

Definition 2.8 $BX : S \times \text{Ags} \times S \mapsto [0, 1]$ is a subjective probability function concerning possible next static states, defined by

$$BX(s, ag, s') = \prod_{ag' \in \text{Ags}} B(s, ag, ag', E(s, h, ag')) \text{ with } s' = \text{succ}(s, h) \text{ for some } h$$

Note that $BX(s, ag, s')$ expresses agent ag 's belief in state s that its choice ends up in s' modulo the assumption that ag actually chooses such as to make s' a possible outcome; if ag chooses such that s' is excluded by its choice, the chance for s' is of course 0.

Now before we can define the notion of ‘seeing to it under a minimum bound on the probability of success’ formally as a truth condition on the frames of Definition 2.7 we need to do more preparations. First we assume that the intersection of the h -effectivity functions of all agents together yields a unique static state. We can safely assume this, because this condition is not modally expressible. This justifies Definition 2.9 below, that establishes a function characterizing the static states next of a given state that satisfy a formula φ relative to the current choice of an agent.

Definition 2.9 The ‘possible next static φ -states’ function $\text{Pos}X : S \times H \times \text{Ags} \times \mathcal{L} \mapsto 2^S$ which for a state s , a history h , an agent ag and a formula φ gives the possible next static states obeying φ given the agent's current choice determined by h , is defined by: $\text{Pos}X(s, h, ag, \varphi) = \{s' \mid s' \in E(s, h, ag) \text{ and } \langle s', h' \rangle \models \varphi \text{ for all } h' \text{ with } s' \in h'\}$.

Now we can formulate the central ‘chance of success’ (CoS) function that will be used in the truth condition for the new operator. The chance of success relative to a formula φ is the sum of the chances the agent assigns to possible next static states validating φ .

Definition 2.10 The chance of success function $\text{Co}S : S \times H \times \text{Ags} \times \mathcal{L} \mapsto [0, 1]$ which for a state s and a history h an agent ag and a formula φ gives the chance the agent's choice relative to h is an action resulting in φ is defined by: $\text{Co}S(s, h, ag, \varphi) = 0$ if $\text{Pos}X(s, h, ag, \varphi) = \emptyset$ or else $\text{Co}S(s, h, ag, \varphi) = \sum_{s' \in \text{Pos}X(s, h, ag, \varphi)} BX(s, ag, s')$.

⁴ I believe however, that there is a glitch in the terminology surrounding the phenomena of dependence in *stit* theory. I now prefer to talk about “independence of choices” and belief this corresponds to “dependence of agency”.

Extending the probabilistic frames of Definition 2.7 to models in the usual way, the truth condition of the new operator is defined as follows.

Definition 2.11 *Relative to a model $\mathcal{M} = \langle S, H, E, B, V \rangle$, truth $\langle s, h \rangle \models [ag \text{ xstit}^{\geq c}] \varphi$ of a formula $[ag \text{ xstit}^{\geq c}] \varphi$ in a dynamic state $\langle s, h \rangle$, with $s \in h$, is defined as:*

$$\langle s, h \rangle \models [ag \text{ xstit}^{\geq c}] \varphi \Leftrightarrow CoS(s, h, ag, \varphi) \geq c$$

Using the example model of Fig. 3 we can now discuss truth evaluations on probabilistic *stit* models. As we saw earlier, relative to static state s_1 and the history h_5 that is part of the bundle of histories Hb_5 we do not have that the choice by agent ag_1 ensures that φ holds, since the other agent has two choices (the bottom one and the top one) for which φ will not be true. So in this model we have that $\langle s_1, h_5 \rangle \not\models [ag_1 \text{ xstit}^{\geq 1}] \varphi$. But we do have that $\langle s_1, h_5 \rangle \models [ag_1 \text{ xstit}^{\geq 0.3}] \varphi$ since ag_1 believes that with a chance of 0.3 agent ag_2 exercises the choice of the middle row. But, relative to histories in, for instance the bundle Hb_1 , agent ag_1 has better chances to see to it that φ will be true. In particular we have that $\langle s_1, h_1 \rangle \models [ag_1 \text{ xstit}^{\geq 0.4}] \varphi$, because it can add up the chances of the bottom two rows. Note that this is also true relative to the histories in bundle Hb_3 for which the result is $\neg \varphi$. Here we have a situation where the agent saw to it that φ with a chance of success of at least 0.4, but failed. Also note that situations like these show that it is consistent in the logic to have, for instance, that $[ag_1 \text{ xstit}^{\geq c}] \varphi \wedge [ag_2 \text{ xstit}^{\geq c}] \neg \varphi$, that is, if c is not 1.

The probabilistic *stit* operator we gave in Definition 2.11 faithfully generalizes the *stit* operator of our base XSTIT^P system: the objective *stit* operator $[ag \text{ xstit}] \varphi$ discussed in Sect. 2.1 comes out as the probabilistic *stit* operator assigning a probability 1 to establishing the effect φ . This is very natural. Where in the standard *stit* setting we can talk about ‘ensuring’ a condition, in the probabilistic setting we can only talk about establishing an effect with a certain lower bound on the probability of succeeding.

We now give a Hilbert system for the probabilistic *stit* logic. The system is parametric in probabilistic variables c and k . This means that the system encodes infinitely many axioms, since there can be infinitely many values for c and k . To obtain a standard Hilbert system we can pose a prior limit to the possible values of probabilities.

Definition 2.12 *Relative to the semantics following from Definitions 2.4 and 2.11 we define the following Hilbert system. We assume all the standard derivation rules for the normal modalities X and \Box . Furthermore, we assume the standard derivation rules for the weak modality $[ag \text{ xstit}^{\geq c}] \varphi$, like closure under logical equivalence.*

- (p) $p \rightarrow \Box p$ for p modality free
S5 for \Box
- (D) $\neg[ag \text{ xstit}^{\geq c}] \perp$ for $c > 0$
- (Triv) $[ag \text{ xstit}^{\geq 0}] \varphi$
- (Lin) $\neg X \neg \varphi \leftrightarrow X \varphi$
- (Sett) $\Box X \varphi \rightarrow [ag \text{ xstit}^{\geq c}] \varphi$
- (XSett) $[ag \text{ xstit}^{\geq 1}] \varphi \rightarrow X \Box \varphi$
- (Min) $[ag \text{ xstit}^{\geq c}] \varphi \rightarrow [ag \text{ xstit}^{\geq k}] \varphi$ for $c \geq k$
- (Add) $[ag \text{ xstit}^{\geq c}] \varphi \wedge [ag \text{ xstit}^{\geq k}] \psi \rightarrow [ag \text{ xstit}^{\geq c+k-1}] (\varphi \wedge \psi)$ for $c + k > 1$
- (Mon) $[ag \text{ xstit}^{\geq c}] (\varphi \wedge \psi) \rightarrow [ag \text{ xstit}^{\geq c}] \varphi$
- (Dep) $\Diamond [ag_1 \text{ xstit}^{\geq c}] \varphi \wedge \dots \wedge \Diamond [ag_n \text{ xstit}^{\geq k}] \psi \rightarrow$
 $\Diamond ([ag_1 \text{ xstit}^{\geq c}] \varphi \wedge \dots \wedge [ag_n \text{ xstit}^{\geq k}] \psi)$ for $Ags = \{ag_1, \dots, ag_n\}$

Proposition 2.3 (Broersen 2011c) *The Hilbert system is sound relative to the semantics.*

Proposition 2.4 (Broersen 2011c) *The Hilbert system reduces to the complete Hilbert system for xstit after substitution of 1 for the parameter c .*

Note that all axioms for xstit have a natural generalization in the above Hilbert system. The most interesting one is agglomeration that generalizes from the standard normal modal logic axiom (Agl) to the set of weak modal scheme's (Add).

3 Modeling the Determination in Action

The second ingredient of moral luck is the determination of an agent. In particular, moral luck can be described as a moral judgement on the difference between the determination of an agent and the indeterminacy of the result of his action. We define two ways in which to represent the determination of an agent's action. In the first sub-section we argue that the lower bounds given in the previous section already express constraints on the determination on the part of the agent. In the second sub-section we discuss the definition of attempt.

3.1 Risk in Action

Operators of the form $[ag \text{ xstit}^{\geq k}] \varphi$ to a limited extent already express determination on the part of an agent ag . They express that agent ag currently performs a choice where it estimates that k is a lower bound to the probability that φ will obtain. Clearly, this does *not* model determination of the agent in the stronger sense of *intentional* action. But, one can say that the formula gives an 'outer constraint' on the determination of the agent from which some information about the agent's

intentions can be abduced. For instance, if the formula $[ag \ xstit^{\geq 0.8}]p$ is true (the agent chooses an action where it believes the chance to achieve p is at least 0.8), while at the same time the formula $\diamond[ag \ xstit^{\geq 0.1}]p$ is true (the agent *could* have chosen an action where it believes the chance to achieve p can be much lower) then we might explain this by abducing that the agent prefers the higher chance to see to it that p . Of course, from this information we cannot deduce that it is the agent's aim to do p ; that would be jumping to conclusions. But what this small example shows is that the formulas of the logic of the previous section can already be used to specify constraints on the agent's determination. In the next section we will take this one step further by modeling the notion of 'attempt'.

3.2 Attempt

We see an attempt for φ as exercising a choice that is maximal in the sense that an agent assigns the highest chance of achieving φ to it (Broersen 2011a). So we aim to model attempt as a comparative notion. This means, that in our formal definition for the attempt operator $[ag \ xatt]\varphi$ that we introduce here, we drop the absolute probabilities. Let us first go back briefly to Fig. 3 to explain the intended semantics of attempt. We have that for agent 1, the right choice is not an attempt for φ , since the left choice has a higher probability (0.4 vs. 0.3) of obtaining φ . So we have that $\langle s_1, h_5 \rangle \models X\varphi \wedge \neg[Ag_1 \ xatt]\varphi$ and $\langle s_1, h_2 \rangle \models [Ag_1 \ xatt]\varphi$. We can also see in the picture that an attempt is not necessarily successful: $\langle s_1, h_3 \rangle \models X\neg\varphi \wedge [Ag_1 \ xatt]\varphi$.

We now give the formal definition. The truth condition for the new operator $[ag \ xatt]\varphi$ is as follows.

Definition 3.1 *Relative to a model $\mathcal{M} = \langle S, H, E, B, \pi \rangle$, truth $\langle s, h \rangle \models [ag \ xatt]\varphi$ of a formula $[ag \ xatt]\varphi$ in a dynamic state $\langle s, h \rangle$, with $s \in h$, is defined as:*

$$\begin{aligned} \langle s, h \rangle \models [ag \ xatt]\varphi &\Leftrightarrow \\ \forall h' : & \text{if } s \in h' \text{ then } CoS(s, h', ag, \varphi) \leq CoS(s, h, ag, \varphi) \\ \text{and} & \\ \exists h'' : & s \in h'' \text{ and } CoS(s, h'', ag, \varphi) < CoS(s, h, ag, \varphi) \end{aligned}$$

This truth condition explicitly defines the comparison of the current choice with other choices possible in that situation. In particular, if and only if the chance of obtaining φ for the current choice is higher than for the other choices possible in the given situation, the current choice is an attempt for φ . The 'side condition' says that there actually must be a choice alternative with a strictly lower chance of success.

Proposition 3.1 (Broersen 2011a) *Each instance of any of the following formula schemas is valid in the logic determined by the semantics of Definition 3.1.*

- (Cons) $\neg[ag \text{ xatt}]\perp$
 (D) $[ag \text{ xatt}]\neg\varphi \rightarrow \neg[ag \text{ xatt}]\varphi$
 (Dep-Att) $\diamond[\{ag1\} \text{ xatt}]\varphi \wedge \diamond[\{ag2\} \text{ xatt}]\psi \rightarrow$
 $\diamond([\{ag1\} \text{ xatt}]\varphi \wedge [\{ag2\} \text{ xatt}]\psi)$
 (Sure-Att) $[ag \text{ xstit}]\varphi \wedge \diamond\neg[ag \text{ xstit}]\varphi \rightarrow$
 $[ag \text{ xatt}]\varphi$

The D-axiom says that the same choice cannot be at the same time an attempt for φ and $\neg\varphi$. This is due to the presence of the ‘side condition’ in Definition 3.1. The side condition says that a choice can only be an attempt if there is at least one alternative choice with a strictly lower chance of success. Now we see immediately why the D-axiom holds: this can never be the case for complementary effects, since these have also complementary probabilities. In *stit* theory, side conditions are used to define ‘deliberative’ versions of *stit* operators (Horty and Belnap 1995). And indeed the same intuition is at work here: a choice can only be an attempt if it is ‘deliberate’.

The (Indep-Att) schema says that attempts of different agents are independent. Attempts are independent, because maximizing choice probabilities from the perspective of one agent is independent from maximizing choice probabilities from the perspective of some other agent.

Finally, the (Sure-Att) schema reveals the relation between the *stit* operator of our base language and the attempt operator. We saw that we can associate the operator $[ag \text{ xstit}]\varphi$ with a probabilistic *stit* operator with a chance of success of 1. Now, if such a choice qualifies as an attempt, it can only be that there is an alternative to the choice with a probability strictly lower than 1 (due to the side condition in Definition 3.1). In the base language we can express this as the side condition $\diamond\neg[ag \text{ xstit}]\varphi$ saying that φ is not ensured by *ag*’s choice. This results in the property (Sure-Att) that says that if *ag* ensures φ with a chance of success of 1, and if *ag* could also have refrained (i.e., *ag* took a chance higher than 0 for $\neg\varphi$), then *ag* attempts φ . This again reveals the relation between the notion of attempt and the notion of ‘deliberate choice’ from the *stit* literature (Horty and Belnap 1995).

4 Moral Obligations, Prohibitions and Luck

The third ingredient of a formalization of both moral luck and legal luck is the normative aspect. In this section we will formalise obligation and prohibition in the moral sense. In the next section we do the same for obligation and prohibition in the legal sense. Adapting the approach put forward in Broersen (2011b) to the case of probabilistic action, we will use Anderson’s reduction of normative truth to logical truth (Anderson 1958) to express either the legal or the moral evaluation of the result of an action against ethical or legal normative codes. We do not explicitly represent these normative codes. However, for future research it will be interesting to investigate how moral luck might depend on the specific moral (ethical) normative code used to evaluate actions. The reduction enables us to express normative assertions

about the good or bad determination in an agent's action. We give four definitions. The first is about being morally forbidden to take a risk.

Definition 4.1 A (moral) prohibition for agent ag to perform a choice by which ag believes to take a risk of at least k to obtain φ , denoted $Forb_{Mor}[ag \text{ xstit}^{\geq k}]\varphi$, is defined by:

$$Forb_{Mor}[ag \text{ xstit}^{\geq k}]\varphi =_{def} \Box([ag \text{ xstit}^{\geq k}]\varphi \rightarrow [ag \text{ xstit}^{\geq 1}]Viol)$$

The definition makes a link between action results in two different realms. The condition φ is an action effect in the physical realm that is subject to the moral prohibition $Forb[ag \text{ xstit}^{\geq k}]\varphi$. The condition $Viol$ can be thought of as an action effect in *social* reality (Searle 1995). The definition defines prohibition by relating the effects in both realities. In the examples below we will discuss why this formalises *moral* prohibition rather than *legal* prohibition. First we define other deontic modalities. The pattern of Definition 4.1 is repeated in the definitions below. The first is about being morally obliged to preserve a given lower bound on the chance on success. The second and third definition are about being morally forbidden and being morally obliged to attempt.

Definition 4.2 A (moral) obligation for agent ag to perform a choice by which ag believes to have a chance of at least k to obtain φ , denoted $Obl_{Mor}[ag \text{ xstit}^{\geq k}]\varphi$, is defined by:

$$Obl_{Mor}[ag \text{ xstit}^{\geq k}]\varphi =_{def} \Box(\neg[ag \text{ xstit}^{\geq k}]\varphi \rightarrow [ag \text{ xstit}^{\geq 1}]Viol)$$

Definition 4.3 A (moral) prohibition for agent ag to attempt φ , denoted $Forb_{Mor}[ag \text{ xatt}]\varphi$, is defined by:

$$Forb_{Mor}[ag \text{ xatt}]\varphi =_{def} \Box([ag \text{ xatt}]\varphi \rightarrow [ag \text{ xstit}^{\geq 1}]Viol)$$

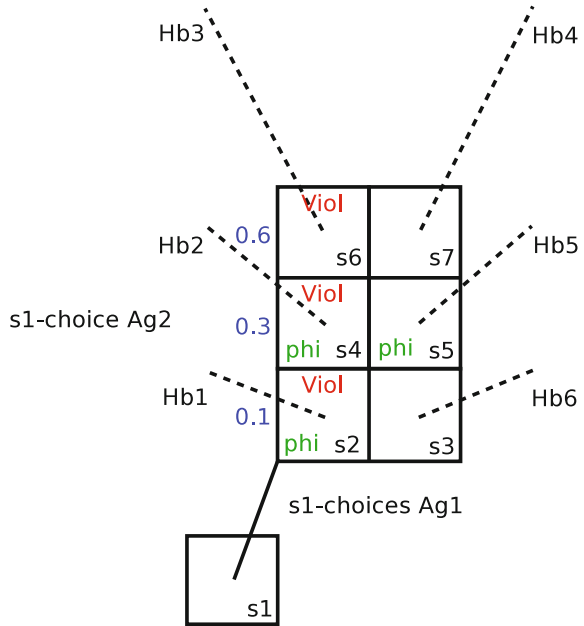
Definition 4.4 A (moral) obligation for agent ag to attempt φ , denoted $Obl_{Mor}[ag \text{ xatt}]\varphi$, is defined by:

$$Obl_{Mor}[ag \text{ xatt}]\varphi =_{def} \Box(\neg[ag \text{ xatt}]\varphi \rightarrow [ag \text{ xstit}^{\geq 1}]Viol)$$

We can also make meaningful variants of the definitions where conditions $\neg[ag \text{ xatt}]\varphi$ are replaced by $[ag \text{ xatt}]\neg\varphi$. The notions resulting are weaker from a normative perspective, since for these variants the agent is only in violation if it explicitly sees to the bad thing happening. For a non-probabilistic setting, nuances like these are explained in Broersen (2011b).

Note that obligations and prohibitions are moment determinate. This means that their truth value is the same for any history through a specific static state. This is due to the presence of the historical necessity operator ' \Box ' as the first operator in all the

Fig. 4 In state s_6 the agent is lucky that $\neg\varphi$

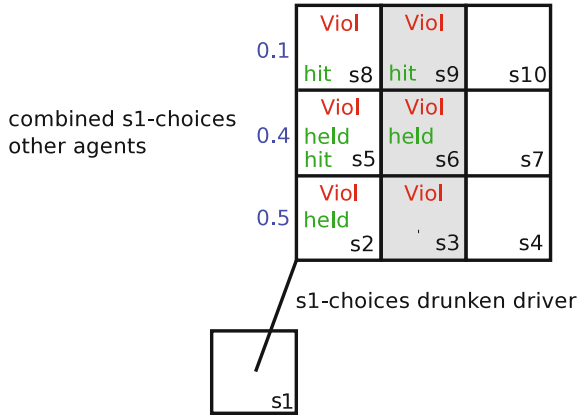


definitions. So it is assumed here that what an agent is obliged or is forbidden does not depend on what it is doing or on what others are doing.

Before discussing the moral character of the prohibitions and obligations defined, let us look at some example formulas. We discuss these normative formulas by interpreting them in the model of Fig. 4, which adds violation constants to the earlier discussed model of Fig. 3. Our first example formula, which is satisfied in dynamic state $\langle s_1, h_3 \rangle$ in Fig. 4 is $Forb[Ag_1 \text{ xstit}^{\geq 0.4}]\varphi \wedge [Ag_1 \text{ xstit}^{\geq 0.4}]\varphi \wedge X\neg\varphi$. The formula expresses that agent Ag_1 is forbidden to choose in such a way that it believes to have a risk of at least 0.4 to obtain φ , while at the same time agent Ag_1 is actually doing such an action, but, where it is lucky since what is actually happening is that $\neg\varphi$ is obtained (state s_6 in the model). The second example formula is very closely related. It is also satisfied in dynamic state $\langle s_1, h_3 \rangle$ of Fig. 4. The formula is $Forb[Ag_1 \text{ xatt}]\varphi \wedge [Ag_1 \text{ xatt}]\varphi \wedge X\neg\varphi$ and it expresses that agent Ag_1 is forbidden to attempt φ , while at the same time that is actually what it is doing, but, where it is lucky since what is actually happening is that $\neg\varphi$ is obtained (again, state s_6 in the model).

So, for both example formulas state s_6 is a state of luck. The condition φ the agent according to the normative part of the formula is supposed to avoid, is indeed not true in it, but that is not due to the determination of the agent, but due to the lucky coincidence that the other agent took the top row as its choice. The agent is lucky in both example formula situations; the first formula specifies how the agent deliberately took a significant risk for φ and the second formula specifies how the agent even explicitly attempted φ . So definitely, the state s_6 is a state of luck. But, it

Fig. 5 Moral violations by the drunken driver



is not of the moral kind. Even though it is justified to say that the agent is lucky, there is still a violation. The violation is due to the fact that the agent exercised the wrong *choice*; one that went against its moral obligations or prohibitions. So, although the outcome is not bad, there is still a violation, representing that the agent was morally wrong. But if that is the interpretation, then in this setting moral luck does not exist. This observation connects to one made by Bernard Williams himself. Williams said (Williams 1993) that when he coined the term ‘moral luck’ he thought it would be an oxymoron. This formalisation represents that initial opinion; agents are *not* morally lucky if the outcome by coincidence is according to the moral obligations and prohibitions, because in that case there is still a *moral violation*.

We are now in the position to come back to the drunken driver example. We discuss the two scenarios of the example by relating them to the model of Fig. 5. First there are several things to explain about this model’s representation of the scenario. In the model, the middle ‘grey’ column choice is the choice taken by the driver; the choice to drive after drinking. The short hand *hit* stands for the driver’s car hitting a person. The short hand *held* stands for the driver being held by the police. Now the first important thing to be explained about the modelling of the scenarios by the pictured model concerns the chosen granularity of the actions. In the model, the exact granularity of the choices is left unspecified. Of course, the choice of the driver to take the risk and drive is not what game theorists would call a ‘one shot’ action; it typically involves several choices in a row. The driving itself, that at least takes place for 1 km in the scenarios, is an action/process with duration that stretches out from the initial moment of choosing to the moment of hitting/being held by the police. It can be that during this period the agent reconsidered but saw no possibility to undo its choice, or it can be that the agent reinstated its decision, being reinforced by the idea that so far everything went smoothly. But all this is abstracted away from. Game theorists would say that the ‘extensive’ game form is normalised to a ‘normal’ game form. Actually, the possibility to look at scenarios more abstractly by normalising the situation is the big advantage of *stit* formalisms. If we would

have to model the same scenario in a dynamic logic (Harel et al. 2000) or situation calculus (McCarthy 1963) formalism, or if we would develop a Davidsonian event-based theory (Davidson 1980), we would have to commit to some bottom level of action/event description. *Stit* theory, on the other hand, allows us to take *any* level of granularity in the description of action. And here we assume exactly the right level of description for the problem at hand.

A second modelling choice to explain are the choices of the other agents in the scenario. There are at least two other agents: a police man and the person risking being hit. In the picture we combine their choices. Actually, we will not be clear about what the exact combined choices of these two other agents are. We see that the model assigns three such choices to this sub-group (the three rows), without making explicit what combined choices they represent. In particular, we do not assign probabilities to the choices of the drunken driver as seen from the perspective of the other two agents. Without these probabilities, we cannot say much about the character of the three row choices. But that is not a problem. The example is about the choices of the drunken driver, and the choices for the drunken driver are clear. There is the right-most choice, which is the choice of avoiding being held by the police and avoiding hitting any person (maybe taking a taxi). There is the middle choice of taking the risk to be held by the police (subjective chance of 0.4) and hitting a person (subjective chance of 0.1). The reason the agent takes this risky choice is that he hopes to end up in state s_3 , that is, in the state where he reaches home without any problems. The left-most choice is one with an optimal chance of either being held by the police or hitting a person, or both. This choice is there in order to make clear that the middle choice is not one that optimises the chances of either hitting a person or being held by the police, that is, the agent does not attempt to hit a person, and does not attempt to be held by the police; it is just being negligent. A more precise model of the situation, with all relevant choices of all three involved agents explicitly represented would be much more extensive. Here we only represent the choices of the drunken driver to discuss the phenomenon of moral luck.

The question raised by the possibility of moral luck is whether or not we should judge occurrence of the situation s_6 and occurrence of the situation s_9 differently. In practice we seem to do that, as the drunken driver example aims to portray. If it is s_6 that comes out of the agent's choice, it is fined 300 Euro, and if it is s_9 that comes out, it is convicted for manslaughter, even though both are the result of exercising the middle 'grey' choice in the figure, which means that the agency involved in both possibilities is exactly the same. And of course, as the model shows, the agent can be even more lucky and end up in state s_3 . This possibility is the reason that the agent exercises this choice in the first place. Then, in state s_6 , the agent is lucky and not lucky at the same time. It is lucky, because it could have ended up in state s_9 which would be worse. But it is also unlucky, because it could have ended up in state s_3 , which would have been much better. But again we can ask the question: what kind of luck are we considering here? If it would be genuine moral luck, and if violations are moral violations, then the modelling is problematic, since our definitions demand that all three states s_3 , s_6 and s_9 are violation states. The solution I want to suggest is that there can be a justification different from a moral justification for the fact that

there is a difference in the legal treatment of the different outcomes. The justification can be found by making clear the purpose of our legal systems. Before I give the argument, we will look at the formalisation of legal prohibition and obligation.

5 Legal Obligations, Prohibitions and Luck

The drunken driver example is built around the phenomenon that the legal evaluation of the different outcome states is likely to be different. Of course, the exact differences depend on the legal normative code relative to which we evaluate the outcomes. But, this will not be of our concern. Instead our concern will be the formalisation of legal obligation and prohibition. For the formal definition of the legal versions of these deontic operators we will have to consider different violation conditions. Figure 6 gives a possible model for such legal violation conditions in the drunken driver example. We see two differences with the model of Fig. 5: first, violations for the decision to drink and drive only occur if indeed the agent is either held by the police or hits a person, and second, the violations for being held and hitting a person are different, which reflects that legal systems evaluate such outcomes differently. We can now adapt Definition 4.1 and the other definitions for moral obligations and prohibitions to the legal context. The formal definitions reflect the dependency on outcomes by introducing an extra condition on effects in the normative realm (which is part of social reality), namely that the effect indeed must have occurred. This would bring us to the following characterisation of the legal prohibition to take a risk.

$$Forb_{Leg}[ag \ xstit^{\geq k}] \varphi = \Box([ag \ xstit^{\geq k}] \varphi \rightarrow [ag \ xstit^{\geq 1}] (\varphi \rightarrow Viol))$$

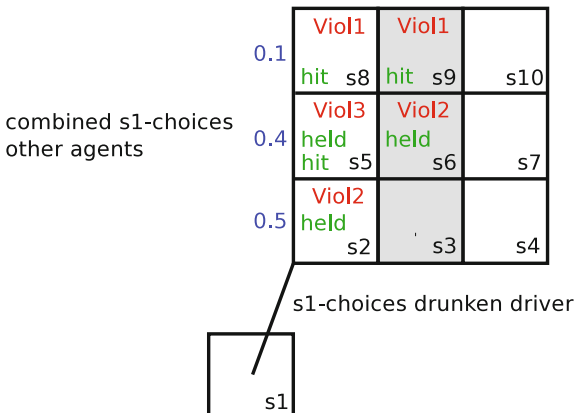
However, we can use the logic to simplify this characterisation. Since the operator $[ag \ xstit^{\geq 1}] \varphi$ is normal (which is not the case for other values of k), it obeys the K axiom. Furthermore, we have that $[ag \ xstit^{\geq 1}] \varphi \rightarrow [ag \ xstit^{\geq k}] \varphi$. Using these properties we can show that the above definition, for any specific instantiation of the propositional meta-variable φ , is equivalent to the characterisation as in the following definition.

Definition 5.1 *The legal prohibition to exercise a choice for which there is a subjective risk of at least k that it has φ as an outcome, denoted $Forb_{Leg}[ag \ xstit^{\geq k}] \varphi$, is defined by:*

$$Forb_{Leg}[ag \ xstit^{\geq k}] \varphi =_{def} \Box([ag \ xstit^{\geq 1}] \varphi \rightarrow [ag \ xstit^{\geq 1}] Viol)$$

Here we see how in a legal definition of prohibition that is strictly based on outcomes, the subjective element, in the context of this prohibition represented by the number k , is eliminated from the definition; the only things that count are if φ

Fig. 6 Possible legal violations by the drunken driver



indeed occurs and if it occurs due to the involvement of agent *ag*. Using a similar line of reasoning we come to the following characterisation of legal obligation.

Definition 5.2 *The legal obligation to exercise a choice for which there is a subjective chance of at least k that it has φ as an outcome, denoted $Obl_{Leg}[ag \ xstit^{\geq k}]\varphi$, is defined by:*

$$Obl_{Leg}[ag \ xstit^{\geq k}]\varphi =_{def} \Box(\neg[ag \ xstit^{\geq 1}]\varphi \rightarrow [ag \ xstit^{\geq 1}]\text{Viol})$$

For coming to definitions of ‘the legal prohibition to attempt’ and ‘the legal obligation to attempt’, we cannot assume a property like $[ag \ xstit^{\geq 1}]\varphi \rightarrow [ag \ xatt]\varphi$. This property does not hold, because an attempt cannot be an attempt if there is no alternative (see Proposition 3.1). This means that we cannot perform the same elimination as for the ‘risk’ versions of the operators, as given above. We come to the following definitions.

Definition 5.3 *The legal prohibition to attempt, denoted $Forb_{Leg}[ag \ xatt]\varphi$, and the legal obligation to attempt, denoted $Obl_{Leg}[ag \ xatt]\varphi$, are defined by:*

$$Forb_{Leg}[ag \ xatt]\varphi =_{def} \Box([ag \ xatt]\varphi \rightarrow [ag \ xstit^{\geq 1}](\varphi \rightarrow \text{Viol}))$$

$$Obl_{Leg}[ag \ xatt]\varphi =_{def} \Box(\neg[ag \ xatt]\varphi \rightarrow [ag \ xstit^{\geq 1}](\neg\varphi \rightarrow \text{Viol}))$$

The definitions do not reflect that violations for different outcomes are different; they only reflect that a violation depends on the occurrence of a bad outcome as such. But different violations for different bad outcomes are easily added to the picture. We can work with separate violation constants for violations of separate prohibitions and obligations.

6 Discussion

The two formalizations I have given, the one of Sect. 4 and the one of Sect. 5 represent two extremes. In the formalisation of moral deontic operators in Sect. 4 violations are the result of the moral evaluation of an agent's subjective choices. If an agent attempts something wrong, or is negligent by taking a risk, it is in violation, independent of the outcome. In the formalisation of Sect. 5 we have the other extreme: the choices themselves are not evaluated, but only their outcomes. And this reflects legal practice; legal systems cannot inspect the subjective considerations accompanying choices of agents and have to rely on outcomes. But, of course, this does not yet explain why in our example hitting a person and being held by the police are evaluated differently by the legal system while there is objective evidence (0.15 % alcohol) that the subjective risk-taking behind both scenarios is the same. There are several possible explanations for this phenomenon. The first is that legal evidence of similarity in risk taking is still 'only' evidence. It does not *proof* with 100 % certainty that the risk-taking in both situations was the same. And this means that there will always be some influence from the actual outcome on the evaluation of the level of risk-taking; we can gather as much evidence about the similarity in risk-taking as we can find, still there will always be the suspicion that in the situation where the outcome was worse, the risk-taking was higher. A second explanation for the phenomenon is that legal systems are not so much directed at the regulation of individual behaviour but at the regulation of societies of agents. Legal systems are not always fair towards individuals; they sacrifice fairness towards the individual to the general benefits of regulation for the society as a whole. It is generally felt that it gives the wrong signal to other possible offenders to let somebody who drank ten beers and fatally hits a person get away with a 300 Euro fine. And it does not make a difference if it is true that if instead this agent would have been held by the police, that same 300 Euro would have been the fine for drinking and driving. Furthermore, if specialists like police investigators, lawyers and judges will have difficulty assessing the similarity in risk taking for the two scenarios, then certainly the general public will. The society as a whole will simply demand higher penalties for outcomes that are less lucky, so that is what the laws of our legal systems reflect. Indeed, this argument ultimately boils down to the observation that our legal systems cannot avoid a certain level of scapegoat justice.

Given the observed differences between the moral and legal evaluation of actions, there is an obvious explanation for the problem introduced by the phenomenon of moral luck: our views about the moral assessment of actions are influenced and obscured by our legal views on the matter. So, if that is true, then moral luck does indeed not exist, and the luck involved in the normative assessment of outcomes is always of the legal kind. The confusions surrounding the concept of moral luck are then due to the influence of our legal views on our views on morality.

7 Conclusion

In this paper we considered a formal approach to the understanding of the problem of moral luck. We had to take three steps: (1) we had to account for indeterminacy of action effects, (2) we had to account for the determination in agentic choice, and (3) we had to define the normative evaluation of action. Luck was described as a difference between an agent's determination (i.e., aspect 2) and the outcome of his action (i.e., aspect 1) in the light of a normative assessment of the situation (i.e., aspect 3). The first result of these efforts is a logic framework where we can reason with moral prohibitions and obligations and legal prohibitions and obligations in the context of risk taking actions. The second result is an explanation for the phenomenon of (the belief in) moral luck. Arguably we could also have found this explanation without the formalisation of the notions involved. But I think the formalisation has helped to arrive at the explanation sooner, and helps to argue for its plausibility in a more convincing way.

Several things are left to investigate. A first thing to do would be to find out what the logics are of the defined operators. But more interesting would be to relate the theory put forward here to legal theories about risk taking. For instance, an agent can go legally wrong if it takes risks a 'normal' person would not take. Here we see that a third form of probability is involved; the probability associated with 'normal' risk taking behaviour. We can then say that three forms of probability are relevant for the theory: objective probabilities, that is, likelihood information about what is objectively going on; subjective probabilities, that is, information in probabilistic form about the risks an acting agent believes to be taking; and 'normal' probabilities, that is, probabilistic information about the risk a normal or average person would be taking (this is best thought of as the probabilistic version of common belief). In particular the relation between the latter two forms is essential for the legal assessment of actions and their outcomes. We will have to leave this interesting subject to future research.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Anderson, A.R. 1958. A reduction of deontic logic to alethic modal logic. *Mind* 67:100–103.
- Broersen, J.M. 2011a. Modeling attempt and action failure in probabilistic *stit* logic. In *Proceedings of twenty-second international joint conference on artificial intelligence (IJCAI 2011)*, ed. Toby Walsh, 792–797.
- Broersen, J.M. 2011b. Deontic epistemic *stit* logic distinguishing modes of *mens rea*. *Journal of Applied Logic* 9(2):127–152.

- Broersen, J.M. 2011c. Probabilistic stit logic. In *Proceedings 11th european conference on symbolic and quantitative approaches to reasoning with uncertainty (ECSQARU 2011)*, vol. 6717 of *Lecture notes in artificial intelligence*, ed. W. Liu, 521–531. Springer.
- Broersen, J.M. 2013. Probabilistic stit logic and its decomposition. *International Journal of Approximate Reasoning* 54:467–477. Elsevier.
- Davidson, D. 1980. *Essays on actions and events*. Oxford: Clarendon Press.
- Emerson, E.A. 1990. Temporal and modal logic. In *Handbook of theoretical computer science, volume B: Formal models and semantics*, ed. J. van Leeuwen, Chap. 14, 996–1072. Amsterdam: Elsevier Science.
- Harel, D., D. Kozen, and Tiuryn, J. 2000. *Dynamic logic*. Cambridge: The MIT Press.
- Horty, J.F., and N.D. Belnap. 1995. The deliberative stit: A study of action, omission, and obligation. *Journal of Philosophical Logic* 24(6):583–644.
- McCarthy, John. 1963. Situations, actions, and causal laws. *Technical report*. Stanford: Stanford University.
- Pauly, Marc. 2002. A modal logic for coalitional power in games. *Journal of Logic and Computation* 12(1):149–166.
- Searle, John. 1995. *The construction of social reality*. New York: The Free Press.
- Thomas, Nagel. 1979. Moral luck. In *Mortal questions*, 24–38. New York: Cambridge University Press.
- Williams, Bernard. 1982. Moral luck. In *Moral luck*, 20–39. Cambridge: Cambridge University Press.
- Williams, Bernard. 1993. Postscript. In *Moral luck*, ed. D. Statman. Albany: State University of New York Press.