INFORMATION AND COMPUTATION 94, 1–28 (1991)

Bisimulation through Probabilistic Testing

KIM G. LARSEN AND ARNE SKOU

Department of Mathematics and Computer Science, Aalborg University Center, 9000 Aalborg, Denmark

We propose a language for testing concurrent processes and examine its strength in terms of the processes that are distinguished by a test. By using probabilistic transition systems as the underlying semantic model, we show how a testing algorithm can distinguish, with a probability arbitrarily close to one, between processes that are not bisimulation equivalent. We also show a similar result (in a slightly stronger form) for a new process relation called $\frac{2}{3}$-bisimulation—which lies strictly between that of simulation and bisimulation. Finally, the ultimately strength of the testing language is shown to identify a new process relation called probabilistic bisimulation—which is strictly stronger than bisimulation.

1. MOTIVATION

Since the appearance of Milner's observational view of process behaviour (Milner, 1980), much work has been devoted to the development of theories that support the specification and verification of parallel systems (in particular, compositional theories have been sought). In spite of the extreme importance of this work, there are several reasons which may justify work on the more practical question of testing an implementation against its specification:

- Verification of large systems is at present too costly and time consuming in most situations.
- Implementations are frequently produced using some ad hoc programming language, which excludes a formal verification.
- Most people do not build their own systems; rather, they buy them from some dealer who obviously will make various claims about the abilities of the product. Normally, the dealer will give no information as to how the system was built, and it is therefore impossible for the buyer to verify whether or not the claims hold. Instead, the buyer will be given some amount of time for testing the system after which he must decide whether to keep it (i.e., he believes the claims) or return it (he does not believe the claims).

0890-5401/91 $3.00

• New insight may be gained by looking at concurrency from an alternative (testing) viewpoint.

We consider a test as the description of an algorithm (guaranteed to terminate within some prescribed amount of time) for how to experiment on a machine (process) equipped with buttons. During the execution of a particular test the experimenter (or observer) tries to press one button at a time as precribed by the algorithm, and each time he notes whether it goes down or not (success or failure).

This view of a test accords with previous work on the subject, but there are some remarkable differences in the way the process is controlled during test execution. De Nicola and Hennessy (1983) modestly assume that the test can proceed as long as success is reported, while Phillips (1986) allows it to continue in case of failure. These two approaches are included in the framework of Abramsky (1987), who furthermore requires the ability to take multiple copies of the process at any stage of the test in order to experiment on one copy at a time. As a special option in this feature he also demands that the process can be forced (by the observer) to enumerate all its possible (non-deterministic) transitions under any button, thereby enabling the test to be exhaustive. This very last option Milner calls "controlling the weather conditions" (Milner, 1981), and it makes it possible to test if two processes are bisimulation equivalent (Park, 1981; Milner, 1983) (or rather, whenever two processes are not bisimulation equivalent, there exists a test distinguishing them).

In our view the ability "to control the weather conditions" in the above sense is in direct conflict with the observational viewpoint of process behaviour: no real system is equipped with such a "weather control" knob. Therefore we consider this feature an unealistic one, and as a consequence rule it out. On the other hand we accept the copying feature because in many situations it can be realized by a simple core dump procedure and also because it is an applied procedure in several kinds of fault tolerant systems.

The "weather control" (or global testing), and the bisimulation notion have been criticized by Bloom, Estrail, and Meyer (1988) from the point of view that they cannot be captured as a trace congruence of any "reasonable" process constructions. We agree that the global testing is not a realistic assumption. However, the conclusion of (Bloom *et al.*, 1988) on the untraceability of bisimulation we consider too strong. Actually we show that any difference between two non-bisimilar processes can be detected (with a high probability) by some test. This result is achieved by applying a finer underlying semantic model for processes, namely that of probabilistic transition systems.

In Section 2 the probabilistic model is presented together with a simple

test language. This induces another language for writing down the experiences that may be observed during an execution of a test, and as a consequence of applying the probabilistic model, each process defines a probability distribution over the observation language for a given test.

The main result of this paper is a systematic framework for testing a process against its specification. We consider a specification as being formed by a number of desired properties of the final implementation. Such properties may be formulated in, e.g., modal logic, temporal logic, or process algebra. In Section 3 we introduce the notion of *testable property* as a property which through testing may correctly be decided to hold of processes with arbitrarily high probability. In the remaining sections the properties of three different modal logics are shown to be testable. One of the logics is the well known Hennessy–Milner Logic (Hennessy and Milner, 1985) and the other two are the restriction of HML called Limited Modal Logic in (Bloom *et al.*, 1988) and a new Probabilistic Modal Logic. For each logic we also give an operational definition of the associated process equivalence. One of these is of course the usual bisimulation relation whereas the finest is a new relation called probabilistic bisimulation.

Proofs are given in the Appendix.

2. Probabilistic Transition Systems. A Simple Test Language

In order to describe a test as an algorithm running on a process, we adopt the well-established notion of labelled transition systems (Plotkin, 1981) as a tool for defining the operational behaviour of processes. This model has been extensively used over the last few years for describing properties of communicating processes and for defining relations under which such processes are to be considered indistinguishable (Pnueli, 1985). Bisimulation is one such relation preserving deadlock properties, and it has been argued by Milner (Milner, 1981) that one has to control the non-determinism (i.e., the weather control) of processes, if non-bisimilarity is going to be detectable by a test. As we find "weather control" unrealistic, the model must be refined instead, and one immediate refinement is to consider each transition as happening with some fixed probability according to the definition below. Thus, even though the observer cannot himself control the weather, he can now—due to the probabilistic nature of transitions—with arbitrarily high degree of confidence assume that all transitions have been examined, simply by repeating the experiment many times (using of course the copying facility).

Definition 2.1. A *probabilistic transition system* is a tuple

$$\mathscr{P} = (\mathrm{Pr}, \mathrm{Act}, \mathrm{Can}, \mu),$$

where Pr is a set of processes (or states), Act is the set of (observable) actions that processes may perform, Can is an Act-indexed family of sets of processes, with Can_a indicating the set of processes that can perform the action a, and μ is a family of probability distributions, $\mu_{p,a}: \text{Pr} \to [0, 1]$, for any $a \in \text{Act}$ and $p \in \text{Can}_a$, indicating the possible next states (and their probabilities) after p has performed a.

Note that whenever $p \in \text{Can}_a$ we have $\sum_{p'} \mu_{p,a}(p') = 1$ since $\mu_{p,a}$ is a probability distribution. Informally, $\mu_{p,a}(p') = \mu$ may be read as "p can perform the action a and with probability μ become the process p' afterwards." Thus p' is a possible next state after a has been performed on p just in case $\mu_{p,a}(p') > 0$. We shall in the remainder of this paper use the following notations:

$$
\begin{aligned}
&p \xrightarrow{a}_{\mu} p' && \text{whenever } p \in \text{Can}_a \text{ and } \mu_{p,a}(p') = \mu \\
&p \xrightarrow{a} p' && \text{whenever } p \xrightarrow{a}_{\mu} p' \text{ for some } \mu > 0 \\
&p \xrightarrow{a} && \text{whenever } p \in \text{Can}_a \\
&p \not\xrightarrow{a} && \text{whenever } p \notin \text{Can}_a.
\end{aligned}
$$

Also, we shall assume that there is a lower limit to the probability of transitions (referred to in the following as the *minimal probability assumption*): i.e., we assume the existence of some $\varepsilon > 0$ such that whenever $p \xrightarrow{a}_{\mu} p'$, then either $\mu = 0$ or $\mu \geqslant \varepsilon$. Clearly this implies that all processes are finitely branching under $\xrightarrow{a}$ for any action a (in fact $\lceil 1/\varepsilon \rceil$ is a universal upper limit on the branching), a condition also known as *image-finiteness* (Hennessy and Milner, 1985). Figure 1 gives examples of probabilistic transition systems.

The execution of a test algorithm on a process basically consists of a series of button pressures (i.e., attempts to observe an action) and as argued in the introduction such an attempt may be issued either on the current process state or on a "fresh" process copy obtained earlier. These basic testing capabilities are reflected in the following simple test language:

DEFINITION 2.2. The *testing language* T has the syntax

$$t ::= \omega \mid a.t \mid (t_1, \ldots, t_n),$$

where ω is a special symbol for termination and $a \in \text{Act}$. If all elements in a tuple test are identical we use the shorthand $(t)^n$.

A test specifies an algorithm for how an observer shall experiment on a process (i.e., which buttons to press when). Informally ω is the test which requires no experiment at all (and therefore will yield no information); $a.t$

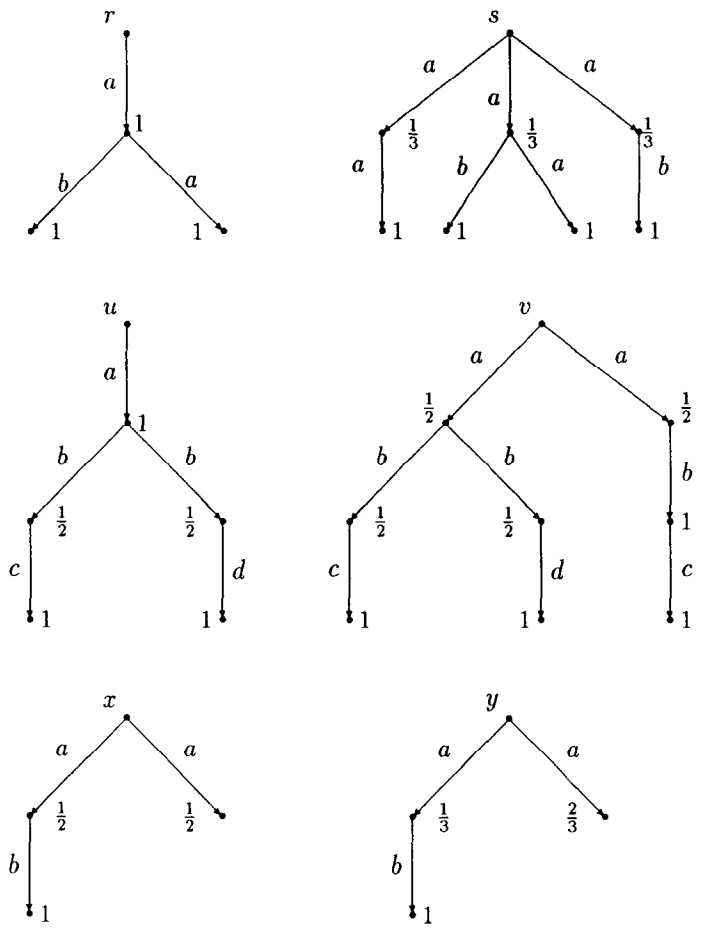


FIG. 1. Probabilistic processes.

describes a test consisting in first applying pressure on the a-button and in case of success proceeding with t. $(t_1, ..., t_n)$ requires that n copies of the current state be taken, allowing all the tests $t_1, ..., t_n$ to be performed independently on the same state. We shall not here give any formal operational semantics of tests, but rather focus on the concept of a *test observation.* During the execution of a particular test on a process, the observer is subject to a series of experiences, each consisting of success or failure of pressing a button. These experiences are written down by the observer according to the (syntactic) structure of the test, and at the end of the execution a full description of all the experiences is available in the form of an *observation.* Obviously—because of the inherent non-determinism of the

processes—there may in general be many possible resulting observations when a test is executed on some process. We define the set of observations that may possibly be obtained from a particular test (on any process) structurally on T as follows:

DEFINITION 2.3. A test t induces the following observation set O_t:

$$O_\omega = \{1_\omega\}$$
$$O_{a.t} = \{0_a\} \cup \{1_a : e \mid e \in O_t\}$$
$$O_{t_1, \ldots, t_n} = O_{t_1} \times \cdots \times O_{t_n}.$$

The observation set for the terminating test ω is just a singleton set. Thus ω provides no basis for distinguishing processes. An observation of a test $a.t$ is either 0_a, indicating that the process did not respond to a (thus the test terminates in this case), or of the form $1_a : e$, where 1_a indicates positive response on a and e is an observation of the following test t. An observation of a tuple test $(t_1, \ldots, t_n)$ consists of tuples of observation $(e_1, \ldots, e_n)$ with $e_i \in O_{t_i}$, and we shall use the shorthand $(e)^n$, when all e_i in a tuple are identical.

Figure 2 contains three tests and their associated observation sets (using a slightly liberal notation for sets, with $1_a : U$ and $\{0_a, U\}$ abbreviating $\{1_a : e \mid \in U\}$ and $\{0_a\} \cup U$, respectively).

The execution of a given test t on a particular process p results in some observation within O_t. However, when p is non-deterministic, there may be many possible resulting observations. In our model the non-determinism within processes is modelled probabilistically. Thus the possible resulting observations will occur with different probabilities. In fact any process p defines a probability distribution $P_{t,p}$ on O_t as follows:

DEFINITION 2.4. Let p be a process and t a test. Then $P_{t,p} : O_t \to [0, 1]$ is the probability distribution defined structurally on t as follows:

1. $P_{\omega, p}(1_\omega) = 1$
2. $P_{a.t,p}(0_a) = \begin{cases} 1 & \text{if } p \not\xrightarrow{a} \\ 0 & \text{otherwise} \end{cases}$

   $P_{a.t,p}(1_a : e) = \begin{cases} 0 & \text{if } p \not\xrightarrow{a} \\ \sum_{p'} \mu_{p,a}(p') \cdot P_{t,p'}(e) & \text{otherwise} \end{cases}$ where $e \in O_t$
3. $P_{(t_1, \ldots, t_n), p}((e_1, \ldots, e_n)) = \prod_i P_{t_i, p}(e_i)$ where $\forall i. e_i \in O_{t_i}$.

It is easily verified that for any p and t, $\sum_{e \in O_t} P_{t,p}(e) = 1$ ensuring that the definition above indeed is that of a probability distribution.

test	*observation set*
$a.b.\omega$	$\{0_a, 1_a : E(b)\}$
$a.(a.\omega, b.\omega)$	$\{0_a, 1_a : E(a) \times E(b)\}$
$a.b.(c.\omega, d.\omega)$	$\{0_a, 1_a : \{0_b, 1_b : E(c) \times E(d)\}\}$

FIG. 2. Tests and associated observation set ($E(x) = \{0_x, 1_x : \{1_\omega\}\}$).

In 2, the probability of observing 0_a clearly depends on whether p can perform a or not. If p can perform a, this observation is impossible, and otherwise it is the only one possible. When p cannot perform a, clearly no observation of the form $1_a : e$ is possible. Otherwise the probability is the sum over all processes, where each process contributes with the probability of observing e, but weighted with the probability of the processes being the next state after having performed a.

In 3, assuming independence of the testing on the n copies of p, the probability of a tuple observation is simply the product of the probabilities of the component observations.

Figure 3 shows for three tests (Fig. 2) the probability of the observations with respect to various processes (Fig. 1). However, we shall in the rest of this paper often refer to the probability of subsets $E \subseteq O_t$, rather than the probability of individual elements, and the following laws are easily verified using the fact that $P_{t,p}$ is a probability distribution.

$t = a.b.\omega$

observation e	$P_{t,x}(e)$	$P_{t,y}(e)$
0_a	0	0
$1_a : 0_b$	$\frac{1}{2}$	$\frac{2}{3}$
$1_a : 1_b : 1_\omega$	$\frac{1}{2}$	$\frac{1}{3}$

$t = a.(a.\omega, b.\omega)$

observation e	$P_{t,r}(e)$	$P_{t,s}(e)$
0_a	0	0
$1_a : (0_a, 0_b)$	0	0
$1_a : (0_a, 1_b)$	0	$\frac{1}{3}$
$1_a : (1_a, 0_b)$	0	$\frac{1}{3}$
$1_a : (1_a, 1_b)$	1	$\frac{1}{3}$

$t = a.b.(c.\omega, d.\omega)$

observation e	$P_{t,u}(e)$	$P_{t,v}(e)$
0_a	0	0
$1_a : 0_b$	0	0
$1_a : 1_b : (0_c, 0_d)$	0	0
$1_a : 1_b : (0_c, 1_d)$	$\frac{1}{2}$	$\frac{1}{4}$
$1_a : 1_b : (1_c, 0_d)$	$\frac{1}{2}$	$\frac{3}{4}$
$1_a : 1_b : (1_c, 1_d)$	0	0

FIG. 3. Tables showing $P_{t,p}(e)$ for various p, t, and $e \in O_t$.

LEMMA 2.5. *Let $P_{t,p}$ be the probability distribution for a given process p and a test t. Then the following laws hold for the probabilities $P_{t,p}(E) = \sum_{e \in E} P_{t,p}(e)$ of subsets $E \subseteq O_t$:*

1. $P_{\omega,p}(\{1_\omega\}) = 1$
2. $P_{a.t,p}(\{0_a\}) = \begin{cases} 1 & \text{if } p \not\xrightarrow{a} \\ 0 & \text{otherwise} \end{cases}$

   $P_{a.t,p}(1_a : E_t) = \begin{cases} 0 & \text{if } p \not\xrightarrow{a} \\ \sum_{p'} \mu_{p,a}(p') \cdot P_{t,p'}(E_t) & \text{otherwise} \end{cases}$ *where* $E_t \subseteq O_t$
3. $P_{(t_1,\ldots,t_n),p}((E_1 \times \cdots \times E_n)) = \prod_i P_{t_i,p}(E_i)$
   where $t = (t_1, \ldots, t_n)$ *and* $\forall i. E_i \subseteq O_{t_i}$
4. $P_{t,p}(\varnothing) = 0$
5. $P_{t,p}(E) = 1 - P_{t,p}(E^c)$ *where* $E^c = \{e \in O_t \mid e \notin E\}$.

3. TESTING PROPERTIES OF PROCESSES

We want to use tests for deciding whether a particular implementation is correct wrt a given specification. This question is of a very practical nature and traditionally there has been a strong distinction between a test and a proof in the sense that it is commonly said (Dijkstra, 1972) that

> *testing can be used to show the presence of bugs, but never to show their absence.*

Although this statement is clearly true, it is also generally believed that the more tests a system passes, the more confidence we may have in the correctness of the system. In fact we may hope that extensive testing can confirm the correctness of a system with arbitrary confidence in the following way:

We view a specification as a number of desired properties of the final implementation. Such properties may be described in a number of ways (e.g., modal logic, temporal logic, process algebra) but the important thing is that a property divides the set of possible implementations in two: those enjoying (satisfying) the property and those not enjoying it.

Thus, in the following a property Φ will simply be a set of processes, and we consider an implementation p as being correct with respect to Φ just in case $p \in \Phi$. We want to settle this question of correctness by executing a test on p, and there *may* be situations where the observation e resulting from such a test decides the question completely, but it is more likely that

the observation leaves the question open. This situation is of course not satisfactory.

However, it will be an improvement if we can find a test t_Φ especially made for Φ, such that certain observations E_Φ (called the set of evidence for Φ) occur with very high probability in case $p \in \Phi$, but with very low probability if $p \notin \Phi$. If such a test exists and the observation e resulting from running t_Φ on p is in E_Φ, then we feel fairly confident that $p \in \Phi$ and similarly that $p \notin \Phi$ if e is not in E_Φ. We consider a property as being *testable* if a test along these lines exists for any desired degree of confidence, and the following definition makes this precise:

DEFINITION 3.1. Let $\mathscr{P} = (\mathrm{Pr}, \mathrm{Act}, \mathrm{Can}, \mu)$ be a probabilistic transition system. Then a property $\Phi \subseteq \mathrm{Pr}$ is *testable* iff for any $\delta > 0$ there exist a test t_Φ and an observation set $E_\Phi \subseteq O_{t_\Phi}$ such that it holds for any process $p \in \mathrm{Pr}$ that

1. Whenever $p \in \Phi$, then $P_{t_\Phi, p}(E_\Phi) \geqslant 1 - \delta$
2. Whenever $p \notin \Phi$, then $P_{t_\Phi, p}(E_\Phi) \leqslant \delta$,

where $P_{t_\Phi, p}(E_\Phi) = \sum_{e \in E_\Phi} P_{t_\Phi, p}(e)$.

We shall refer to δ as the *level of significance*, and intuitively it gives an upper bound of making a wrong decision, i.e., concluding that p enjoys Φ, when it does not, or dually concluding that p does not satisfy Φ when in fact it does.

To relate our terminology with the one normally used within hypothesis testing in statistics (Cox and Hinkley, 1974), we may consider the property Φ as a *null hypothesis* H_0, the complementary property Φ^c as the *alternative hypothesis*, and the complement (E_Φ^c) of E_Φ as the *critical region*. The upper limit for $P_{t,p}(E_\Phi^c)$ given that $p \in \Phi$ (which we refer to as the level of significance) is called the *size* of a statistical test (E_Φ, t), and similarly the lower limit of $P_{t,p}(E_\Phi^c)$ given that $p \notin \Phi$ is called the *power* of (E_Φ, t). We shall not here continue the analogy any further, but just mention the strong resemblance between the second requirement of our definition of testability and what might be called *consistency* of a test within statistics.

In the same way we consider a class (collection) of properties as being testable if all its members are testable:

DEFINITION 3.2. A class $\mathscr{C}$ of properties is testable iff for all $\Phi \in \mathscr{C}$, Φ is testable.

The crucial question is then what classes are testable in the above sense and also given a testable class, which processes are equivalent in the sense that no property in the class separates them.

In the remainder of this paper we consider three property classes defined in terms of various modal logics. All three classes are shown to be testable in the above sense, and the associated equivalences on processes are given operational characterizations.

The first and weakest property is given by the restriction of Hennessy–Milner Mogic (HML) called *Limited Modal Logic* in (Bloom *et al.*, 1988). It is shown that two processes are indistinguishable with respect to properties in this class, just in case they define the same sets of possible observations for all tests. Operationally, the equivalence is characterized by a new notion of $\frac{2}{3}$-*bisimulation*, a relation between processes lying strictly between those of simulation (Larsen, 1986) and bisimulation (Park, 1981; Milner, 1983).

The second class of properties, which we show to be testable, is that given by Hennessy–Milner Logic (Hennessy and Milner, 1985). This shows—in contrast to the conclusion reached by (Bloom *et al.*, 1988)—that non-bisimilar processes can indeed be distinguished "at the terminal," at least when our finer probabilistic model is adopted. The operational characterization of HML is of course obtained through *bisimulation* as shown in (Hennessy and Milner, 1985).

The last and strongest class of properties is given by a *Probabilistic Modal Logic* (PML) with the two (next-state) modalities $\langle a \rangle$ and $[a]$ of HML having been replaced by a continuum of (next-state) modalities $\langle a \rangle_\mu$, where $0 \leqslant \mu \leqslant 1$. It is shown that two processes are indistinguishable under PML, just in case they give the exact same probability distribution on the observation set of *any* test. Thus, if two processes are indistinguishable with respect to PML, they are in fact indistinguishable with respect to *any* testable property, and could be called *test equivalent*. Operationally, the equivalence is characterized by a notion of *probabilistic bisimulation*.

4. Limited Modal Logic and $\frac{2}{3}$-Bisimulation

First let us recall the well-known Hennessy–Milner Logic introduced in (Hennessy and Milner, 1985).

Definition 4.1. The formulas of HML are given by the following syntax:

$$F ::= \mathrm{tt} \mid \mathrm{ff} \mid [a]F \mid \langle a \rangle F \mid F_1 \wedge F_2 \mid F_1 \vee F_2.$$

The satisfaction relation, $p \models F$, between processes and formulas is defined as usual for modal logics and Kripke models (see (Hughes and Cresswell,

1972)). Thus, $p \models \langle a \rangle F$, whenever $p' \models F$ for some p' with $p \xrightarrow{a} p'$, and dually, $p \models [a]F$, whenever $p \xrightarrow{a} p'$ implies $p' \models F$.

We shall view any HML formula F as the property (i.e., a set of processes), consisting of all the processes satisfying it.

The fundamental result of (Hennessy and Milner, 1985) is that two processes will satisfy exactly the same HML formulas just in case they are bisimilar (provided $\rightarrow$ is image-finite).

DEFINITION 4.2. Let $\mathscr{P} = (\text{Pr}, \text{Act}, \text{Can}, \mu)$ be a probabilistic transition system. Then *a bisimulation* $\mathscr{R}$ is a binary relation on Pr such that whenever pRq and $a \in \text{Act}$ then the following hold:

1. Whenever $p \xrightarrow{a} p'$, then $q \xrightarrow{a} q'$ such that $p' \mathscr{R} q'$
2. Whenever $q \xrightarrow{a} q'$, then $p \xrightarrow{a} p'$ such that $p' \mathscr{R} q'$.

Two processes p and q are said to be *bisimilar* in case (p, q) is contained in some bisimulation R. We write $p \sim q$ in this case.

Before proving the testability of HML itself, let us consider the restriction, Limited Modal Logic (LML), introduced in (Bloom *et al.*, 1988).

DEFINITION 4.3. The formulas of LML are given by the following syntax:

$$F ::= \text{tt} \mid [a]\,\text{ff} \mid \langle a \rangle F \mid F_1 \wedge F_2.$$

The main limitation of LML is that the formulas only allow a very restrictive use of $[a]$, just enabling the logic to express deadlock on actions. Furthermore we omit the disjunction connective. However, it is easily proved that the addition of $\vee$ to LML does not increase its distinguishing power.

The operational characterization of LML is obtained by the following notion of $\frac{2}{3}$-bisimulation:

DEFINITION 4.4. Let $\mathscr{P} = (\text{Pr}, \text{Act}, \text{Can}, \mu)$ be a probabilistic transition system. Then a $\frac{2}{3}$-bisimulation $\mathscr{R}$ is a binary relation on Pr such that whenever $p \mathscr{R} q$ and $a \in \text{Act}$ then the following hold:

1. Whenever $p \xrightarrow{a} p'$, then $q \xrightarrow{a} q'$ for some q' with $p' \mathscr{R} q'$
2. Whenever $q \xrightarrow{a}$, then $p \xrightarrow{a}$.

Two processes p and q are said to be $\frac{2}{3}$-bisimilar just in case there are $\frac{2}{3}$-bisimulations $\mathscr{R}_1, \mathscr{R}_2$ such that $p \mathscr{R}_1 q$ and $q \mathscr{R}_2 p$. We write $p \simeq q$ in this case.

From the above definition it is clear that the notion of bisimulation is strictly stronger than that of $\frac{2}{3}$-bisimulation (in Fig. 1, $u \simeq v$ but $u \not\sim v$). Also, the notion of $\frac{2}{3}$-bisimulation is strictly stronger than that of simulation (being "half" of bisimulation, cf. (Larsen, 1986)) (in Fig. 1, $r \leqslant s$ and $s \leqslant r$ but $s \not\simeq r$).

Now, LML characterizes $\frac{2}{3}$-bisimulation in the following sense:

THEOREM 4.5. *Let $\mathscr{P} = (\mathrm{Pr}, \mathrm{Act}, \mathrm{Can}, \mu)$ be a probabilistic transition system satisfying the minimal probability assumption. Then two processes are $\frac{2}{3}$-bisimilar just in case they satisfy exactly the same LML formulas.*

In Fig. 1, we see that $s \models \langle a \rangle [b]\,\mathrm{ff}$, whereas $r \not\models \langle a \rangle [b]\,\mathrm{ff}$. Thus, it follows from the above theorem that $s \not\simeq r$.

In order to establish the testability of LML, we introduce properties of the form $[t, e]$ (t being a test and $e \in O_t$), consisting of the processes p for which $P_{t,p}(e) > 0$. Thus p will satisfy $[t, e]$ iff e is a possible resulting observation when executing t on p.

Now, the key fact which will lead to the testability of LML, is that any property described as an LML formula may alternatively be described on the form $[t, e]$, and vice versa.

LEMMA 4.6. (A) *Let F be a LML formula. Then there exist a test τ_F and an observation $e_F \in O_{\tau_F}$ such that $p \models F$ iff $p \in [\tau_F, e_F]$.*

(B) *Let t be a test and e an observation of t. Then there exists a formula $F_{t,e}$ such that $p \models F_{t,e}$ iff $p \in [t, e]$.*

Proof. We give only the constructive definitions of $[\tau_F, e_F]$ and $F_{t,e}$.

(A) $[\tau_F, e_F]$ is defined by structure on F as follows:

$$[\tau_{\mathrm{tt}}, e_{\mathrm{tt}}] = [\omega, 1_\omega]$$

$$[\tau_{\langle a \rangle F}, e_{\langle a \rangle F}] = [a.\tau_F, 1_a : e_F]$$

$$[\tau_{[a]\mathrm{ff}}, e_{[a]\mathrm{ff}}] = [a.\omega, 0_a]$$

$$[\tau_{F_1 \wedge F_2}, e_{F_1 \wedge F_2}] = [(\tau_{F_1}, \tau_{F_2}), (e_{F_1}, e_{F_2})]$$

(B) $F_{t,e}$ is defined by structure on t as follows:

$$F_{\omega, 1_\omega} = \mathrm{tt}$$

$$F_{a.t, 0_a} = [a]\,\mathrm{ff}$$

$$F_{a.t, 1_a : e} = \langle a \rangle F_{t,e}$$

$$F_{(t_1, \ldots, t_n)} = \bigwedge_i F_{t_i} \quad \blacksquare$$

From the above lemma and the characterization theorem 4.5, it follows that two (image-finite) processes are $\frac{2}{3}$-bisimilar just in case they assign non-zero probabilities to exactly the same observations of any test. Thus, it follows from the tables of Fig. 3 that $r \not\simeq s$.

Now, let F be a property described in LML and let $[\tau_F, e_F]$ be its alternative description according to Lemma 4.6. We may then test the property F in the following way: the test t_F will be of the form $(\tau_F)^N$ (where N is to be determined by the desired level of significance), and the evidence set E_F consists of all observations $(e_1, ..., e_N)$ with e_F as a component. By increasing N, clearly the probability of obtaining an obseration within E_F —when indeed $p \models F$—can be made arbitrarily high. On the other hand, the probability of obtaining an observation within e_F will—regardless of N—be 0 when $p \not\models F$.

Under the minimal probability assumption, it is easily verified that in fact $P_{\tau_F, p}(e_F) \geqslant \varepsilon^{|F|}$ whenever $p \models F$, where $|F|$ is the "size" of F (defined inductively by $|\mathrm{tt}| = |[a]\,\mathrm{ff}| = 0$, $|\langle a\rangle F| = 1 + |F|$, and $|F_1 \wedge F_2| = |F_1| + |F_2|$).

Thus, for a given N, $P_{t_F, p}(E_F) \geqslant 1 - (1 - \varepsilon^{|F|})^N$. Hence, in order to meet some specified level of significance δ, we may simply choose N so that $(1 - e^{|F|})^N \leqslant \delta$.

We can now assert the testability of LML in a slightly stronger form:

THEOREM 4.7. *Let F be a LML formula and $\delta > 0$ a desired level of significance. Then there exist a test t_F and an evidence set $E_F \subseteq O_{t_F}$ such that*

1. *Whenever $p \models F$, then $P_{t_F, p}(E_F) \geqslant 1 - \delta$*
2. *Whenever $p \not\models F$, then $P_{t_F, p}(E_F) = 0$.*

Note that when testing for LML properties, we will never conclude that a process satisfies a property when in fact it does not.

EXAMPLE 4.8. Let $F = \langle a\rangle[b]\,\mathrm{ff}$, the level of significance $\delta = 0.1$, and $\varepsilon = \frac{1}{3}$. Then we obtain from the construction of part A in Lemma 4.6 the pair $[\tau_F, e_F] = [a.b.\omega, 1_a : 0_b]$. Now, $(1 - \varepsilon^{|F|})^N \leqslant \delta$ holds for $N \geqslant 6$, and we therefore apply the test $t_F = (\tau_F)^6$ and the set of evidence $E_F = \{(e_1, ..., e_6) \mid \exists e_i = 1_a : 0_b\}$ for a test of F at the level of significance $\delta = 0.1$. For the processes s, r in Fig. 1 we obtain $P_{t_F, s}(E_F) = 1 - P_{t_F, s}(E_F^c) = 1 - P_{t_F, s}((1_a : 1_b : 1_\omega)^6) = 1 - (\frac{2}{3})^6 \approx 0.91$, whereas $P_{t_F, r}(E_F) = 0$. Hence, the test t_F succeeds on s with probability 0.91 and will always fail on r.

5. Hennessy–Milner Logic and Bisimulation

Having established the testability of LML, we now face the full class of HML formulas.

The question of testability of HML boils down to showing how to test $[a]F$ formulas in general. Assume (inductively) that t_F is a test for F with evidence set E_F (with respect to some level of significance δ_F); then $a.t_F$ seems like a suitable basis for testing $[a]F$. However, $a.t_F$ itself will only examine a single a-transition of processes, and the chance of erroneously concluding that a process satisfies $[a]F$, when in fact it does not, may consequently be high. In order to make our conclusion more safe, we must repeat the test $a.t_F$, so that we can assume with high probability that all a-transitions have been examined.

Thus, to test for the property $[a]F$, we propose a test of the form $t_{[a]F} = (a.t_F)^N$, where N is to be determined by the desired level of significance. The set of evidence $E_{[a]F}$ consists of $(0_a, ..., 0_a)$ (indicating that the process cannot perform a, and therefore satisfies $[a]F$) together with all observations $(1_a : e_1, ..., 1_a : e_N)$, where *all* e_i confirm F (i.e., $e_i \in E_F$). With this choice of $t_{[a]F}$ and $E_{[a]F}$, the probability of getting evidence for non-satisfying processes will celarly approach 0 for $N \to \infty$. To illustrate this fact, consider the process p of Fig. 1, and let $F = [a]\langle b\rangle \mathrm{tt}$, a property clearly *not* satisfied by p. Then for $t_F = (a.b.\omega)^N$, $P_{t_F, p}(E_F) = (\frac{1}{2})^N \to 0$ as $N \to \infty$.

Now, let us turn to the problem of testing $\langle a\rangle F$ formulas. Assume (inductively) that t_F is a test for F with evidence set E_F (with respect to some level of significance δ_F). Also in this case, $a.t_F$ seems like a suitable test candidate. However, since $a.t_F$ only examines a single a-transition of processes, we may now erroneously conclude that a process does not satisfy $\langle a\rangle F$ when in fact it does, simply because a "wrong" a-transition was examined. Again, repeating $a.t_F$ will resolve the problem.

Thus, to test a property $\langle a\rangle F$, we use a test of the form $t_{\langle a\rangle F} = (a.t_F)^N$, where N has to be determined from the desired level ofsignificance. However, in this case (in contrast to that of $[a]F$) the set of evidence $E_{\langle a\rangle F}$ consists of the observations $(1_a : e_1, ..., 1_a : e_N)$ with *at least one* e_i confirming F (i.e., $e_i \in E_F$). It is clear that these choices will make the probability of getting evidence for satisfying processes approach 1 as N increases. *Unfortunately*, so does the probability of getting evidence for *non-satisfying* processes. For p not satisfying $\langle a\rangle F$ it is easily shown that $P_{t_{\langle a\rangle F}, p}(E_{\langle a\rangle F})$ is bounded above by $1-(1-\delta_F)^N$. However, $1-(1-\delta_F)^N \to 1$ as $N \to \infty$, so this upper bound does not establish the testability of $\langle a\rangle F$. The way we resolve this discrepancy in the proof is to reverse the order in which δ_F and N are determined. Thus, let the desired level of significance δ be given. Then—assuming $\delta_F = \frac{1}{2}$ say—we first choose N so that

$P_{t_{\langle a\rangle F},p}(E_{\langle a\rangle F}) \geqslant 1-\delta$, whenever p satisfies $\langle a\rangle F$. Having determined N, we now fix δ_F so that when p does not satisfy $\langle a\rangle F$, $P_{t_{\langle a\rangle F},p}(E_{\langle a\rangle F}) \leqslant 1-(1-\delta_F)^N \leqslant \delta$. Clearly this may be done as $(1-\delta_F)^N \to 1$ for $\delta_F \to 0$. It is easy to see that this lowering of δ_F does not decrease $P_{t_{\langle a\rangle F},p}(E_{\langle a\rangle F})$ for p satisfying $\langle a\rangle F$. Based on these informal arguments we now state the following main theorem:

THEOREM 5.1. *The formulas of Hennessy–Milner Logic are testable.*

EXAMPLE 5.2. Consider the processes u and v of Fig. 1. Clearly $u \not\sim v$, and in fact $F = \langle a\rangle[b](\langle c\rangle \mathrm{tt} \wedge [d]\mathrm{ff})$ is a distinguishing property (satisfied by v but not u). The test for F is of the form $t = (a.(b.(c.\omega, d.\omega))^{N_2})^{N_1}$ with the evidence set E containing tuples $(e_1, \ldots, e_{N_1})$ with $e_i = 1_a : (1_b : (1_c : 1_\omega, 0_d))^{N_2}$ or $e_i = 1_a : (0_b)^{N_2}$ for some component e_i. Taking $\delta = 0.2$ (a very modest level of significance), and $\varepsilon = \frac{1}{2}$, the constructions in the proof of Theorem 5.1 yields $N_1 = 6$ and $N_2 = 11$. In order to calculate the actual probabilities for evidence, we introduce the following abbreviations:

$$t = (a.t_1)^{N_1}$$

$$t_1 = (b.t_2)^{N_2}$$

$$t_2 = (c.\omega, d.\omega)$$

$$E_2 = \{(1_c : 1_\omega, 0_d)\} \text{ (for } \langle c\rangle tt \wedge [d]\,\mathrm{ff})$$

$$E_1 = (1_b : E_2)^{N_2} \cup (0_b)^{N_2}$$

$$E = \{(1_a : E_1^c)^{N_1} \cup (0_a)^{N_1}\}^c.$$

Furthermore we enumerate the derivatives of u and v from left to right in a breadth-first manner, and the probabilities of evidence for u and v can now be calculated, using the laws from Lemma 2.5:

$$\begin{aligned} P_{t,v}(E) &= 1 - P_{t,v}(E^c) \\ &= 1 - P_{v,t}((1_a : E_1^c)^{N_1} \cup (0_a)^{N_1}) \\ &= 1 - (P_{t,v}(1_a : E_1^c))^{N_1} \\ &= 1 - (\tfrac{1}{2} \cdot P_{t_1,v_1}(E_1^c) + \tfrac{1}{2} \cdot P_{t_2,v_2}(E_1^c))^{N_1} \end{aligned}$$

$$
\begin{aligned}
P_{t_1,v_1}(E_1^c) &= 1 - P_{t_1,v_1}(E_1) \\
&= 1 - P_{t_1,v_1}((1_b : E_2)^{N_2} \cup (0_b)^{N_2}) \\
&= 1 - (P_{t_1,v_1}(1_b : E_2))^{N_2} \\
&= 1 - (\tfrac{1}{2} \cdot P_{t_2,v_3}(E_2) + \tfrac{1}{2} \cdot P_{t_2,v_4}(E_2))^{N_2} \\
&= 1 - (\tfrac{1}{2} \cdot 1 + \tfrac{1}{2} \cdot 0)^{N_2} \\
&= 1 - \tfrac{1}{2}^{N_2} \\
P_{t_1,v_2}(E_1^c) &= 0.
\end{aligned}
$$

Substituting these values we get for process v

$$
\begin{aligned}
P_{t,v}(E) &= 1 - (\tfrac{1}{2} \cdot (1 - \tfrac{1}{2}^{N_2}))^{N_1} \\
&= 1 - (\tfrac{1}{2} \cdot (1 - \tfrac{1}{2}^{11}))^{6} \\
&\approx 0.984.
\end{aligned}
$$

Likewise we have for process u

$$
\begin{aligned}
P_{t,u}(E) &= 1 - (P_{t,u}(1_a : E_1^c))^{N_1} \\
&= 1 - (1 - P_{t_1,u_1}(E_1))^{N_1} \\
&= 1 - (1 - (P_{t_1,u_1}(1_b : E_2))^{N_2})^{N_1} \\
&= 1 - (1 - (\tfrac{1}{2} \cdot P_{t_2,u_2}(E_2) + \tfrac{1}{2} \cdot P_{t_2,u_3}(E_2))^{N_2})^{N_1} \\
&= 1 - (1 - \tfrac{1}{2}^{N_2})^{N_1} \\
&\approx 0.003.
\end{aligned}
$$

6. Probabilistic Modal Logic and Probabilistic Bisimulation

We now address the strength of our test language in terms of the processes that may be distinguished by some testable property. The previous two sections show that the test language is at least strong enough to distinguish non-bisimilar processes. However, even bisimilar processes may be distinguished. As an example consider executing a test of the form $t = (a.b.\omega)^N$ on the processes x and y of Fig. 1 (which are clearly bisimilar). In the resulting observation, $e = (e_1, ..., e_N)$, we expect the number of occurrences of $1_a : 1_b : 1_\omega$ as a component to be approximately $\frac{1}{2}N$ in the case of x, and $\frac{1}{3}N$ in the case of y. Also—a consequence of Chebyshev's inequality (cf. (Chung, 1974))—the derivations from these

expectations will decrease as N increases. Thus, it seems that we are indeed able to distinguish x and y by a test of the above form.

Of course, this should come as no surprise, as HML and the induced bisimulation equivalence only take into consideration the *mere possibility* of transitions, and abstract away from the actual *probability* with which a possible transition can take place. What is needed in order to characterize precisely the strength of our test language seems to be probabilistic versions of HML and bisimulation. Thus, we propose below a Probabilistic Modal Logic (PML) with the $\langle a \rangle$ and $[a]$ modalities of HML being replaced by a continuum of modalities of the form $\langle a \rangle_\mu$, where a is an action and μ a probability.

DEFINITION 6.1. The formulas of PML are given by the syntax

$$F ::= \mathrm{tt} \mid \mathrm{ff} \mid \Delta_a \mid F_1 \wedge F_2 \mid F_1 \vee F_2 \mid \langle a \rangle_\mu F,$$

where $a \in \mathrm{Act}$ and $\mu \in [0, 1]$.

The *satisfaction relation*, $p \models F$, between processes and formulas of PML is defined as usual for tt, ff, $F_1 \wedge F_2$, and $F_1 \vee F_2$. $p \models \Delta_a$ holds whenever $p \not\xrightarrow{a}$. Hence Δ_a corresponds to the formula $[a]\,\mathrm{ff}$ of HML. Now, we extend the $\longrightarrow$-notation in the following straightforward manner: for $S \subseteq \mathrm{Pr}$ we write $p \xrightarrow{a}_\mu S$ whenever $p \xrightarrow{a}$ and $\sum_{q \in S} \mu_q = \mu$ where $\mu_q = \mu_{p,a}(q)$. Thus, we have as special cases $p \xrightarrow{a}_0 \varnothing$ and $p \xrightarrow{a}_1 \mathrm{Pr}$. Also, $p \xrightarrow{a} S$ will abbreviate $p \xrightarrow{a}_\mu S$ for some $\mu > 0$. Then, we define $p \models \langle a \rangle_\mu F$ whenever $p \xrightarrow{a}_\nu S$ with $\nu \geqslant \mu$ and $\forall q \in S.\, q \models F$, for some S.

From this definition it should be clear that $x \models \langle a \rangle_{1/2} \langle b \rangle_1 \mathrm{tt}$ whereas $y \not\models \langle a \rangle_{1/2} \langle b \rangle_1 \mathrm{tt}$ (see Fig. 1).

Clearly, $\langle a \rangle_\nu F \Rightarrow \langle a \rangle_\mu F$ whenever $\nu \geqslant \mu$. Considering the modalities of HML, the following equivalences hold:

$$\langle a \rangle F \equiv \exists \mu > 0.\, \langle a \rangle_\mu F$$

$$[a] F \equiv \langle a \rangle_1 F \vee \Delta_a.$$

In case the minimal probability assumption holds, we may express $\langle a \rangle F$ directly as follows:

$$\langle a \rangle F \equiv \langle a \rangle_\varepsilon F.$$

Now we face the problem of testing PML formulas, and in particular formulas of the form $\langle a \rangle_\mu F$. Here, we give only an informal account: As for $\langle a \rangle$- and $[a]$-formulas of HML, the test for $\langle a \rangle_\mu F$ will have the form $(a.t_F)^N$, where t_F is a test for F, and N has to be determined by the desired

level of significance. For $\langle a \rangle$-formulas respectively $[a]$-formulas the set of evidence was the tuples with *at least one* respectively *all* components confirming F. For the PML formula, $\langle a \rangle_\mu F$, the evidence set consists of all tuples, $(e_1, ..., e_N)$, where *at least* μN components are of the form $1_a : e_i'$ with e_i' confirming F. However, for technical reasons we have to assume that the transition probabilities cannot come arbitrarily close to each other. We call this requirement the *minimal deviation assumption*, and formaly it means that any value $\mu_{p,a}(p')$ is a multiple of some minimum probability value ε. Assuming that this requirement is fulfilled, the probabilities for evidence will—for increasing N—approach 0 respectively 1 for non-satisfying respectively satisfying processes (using Chebyshev's inequality (Chung, 1974)). Formalizing the above yields the following important result:

THEOREM 6.2. *The formulas of Probabilistic Modal Logic are testable if the minimal deviation assumption is satisfied.*

To obtain an operational account of PML we refine the notion of bisimulation so that probabilities of transitions are catered for.

Whenever $\equiv$ is an equivalence on processes, we write $\mathrm{Pr}/\equiv$ for the set of equivalence classes under $\equiv$. Using this notation, we may formulate the notion of bisimulation equivalence in the following alternative way:

$$p \sim q \Leftrightarrow \forall a \in \mathrm{Act}. \forall S \in \mathrm{Pr}/\sim . \, p \xrightarrow{a} S \Leftrightarrow q \xrightarrow{a} S.$$

Based on the formulation above we then obtain the notion of probabilistic bisimulation as follows:

DEFINITION 6.3. Let $\mathscr{P} = (\mathrm{Pr}, \mathrm{Act}, \mathrm{Can}, \mu)$ be a probabilistic transition system. Then a *probabilistic bisimulation* $\equiv$ is an equivalence on Pr such that whenever $p \equiv q$, then the following holds:

$$\forall a \in \mathrm{Act}. \forall S \in \mathrm{Pr}/\equiv . \, p \xrightarrow{a}_\mu S \Leftrightarrow q \xrightarrow{a}_\mu S.$$

Two processes p and q are said to be *probabilistically bisimilar* in case (p, q) is contained in some probabilistic bisimulation. We write $p \equiv_P q$ in this case.

Figure 4 shows two processes p_1 and p_2 and a probabilistic bisimulation (or rather its equivalence classes) establishing $p_1 \equiv_P p_2$.

As indicated above, PML characterizes probabilistic bisimulation in the following sense:

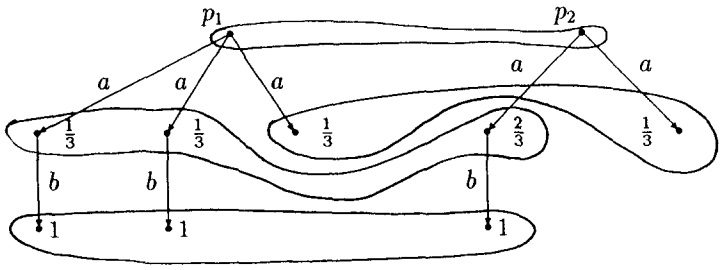


FIG. 4. A probabilistic bisimulation.

THEOREM 6.4. *Let* $\mathscr{P} = (\mathrm{Pr}, \mathrm{Act}, \mathrm{Can}, \mu)$ *be a probabilistic transition system satisfying the minimal deviation assumption. Then two processes are probabilistically bisimilar just in case they satisfy exactly the same PML formulas.*

Even more importantly, it turns out that the notion of probabilistic bisimilarity captures exactly the *limit* as to the distinguishing power of our test language: if two processes are probabilistic bisimilar then *no* testable property will separate them. In fact, we can prove an even stronger claim; namely, that two processes are probabilistic bisimilar just in case they yield the exact same probability distribution on the observation set of *any* test:

THEOREM 6.5. *Let* $\mathscr{P} = (\mathrm{Pr}, \mathrm{Act}, \mathrm{Can}, \mu)$ *be a probabilistic transition system satisfying the minimal deviation assumption. Then* $p \equiv_P q$ *just in case* $P_{t,p}(e) = P_{t,q}(e)$ *for all tests* t *and observations* $e \in \mathscr{O}_t$.

7. CONCLUDING REMARKS AND FUTURE WORK

We have presented a test language and a notion of (probabilistic) testability of process properties. In particular, we have demonstrated that properties expressed as formulas within Hennessy–Milner Logic are testable, and as a consequence that we may distinguish non-bisimilar processes through testing. A Probabilistic Modal Logic has been introduced, and we have shown that the induced notion of probabilistic bisimilarity characterizes the limit as to the distinguishing power of our test language.

The testability results are based on the assumptions of a minimal probability and a minimal deviation respectively. Intuitively the minimum probability assumption is necessary if a process property is going to be testable within a prescribed amount of time, while the minimum deviation assumption may seem a bit technical. However if one wants to test for a

property of form $\langle a \rangle_\mu F$ without this assumption, the following adjustments can be made:

- If it is essential that the applied test not accept any process enjoying $\langle a \rangle_\nu F$, where ν is strictly less than μ, one can alternatively test for the property $\langle a \rangle_{\mu+\varepsilon} F$ (ε being the minimal probability). Of course this has the drawback that processes enjoying $\langle a \rangle_\mu F$ may not pass the test.
- Another refinement is to choose a smaller minimal probability ε. This restricts the class of processes that may pass the test without enjoying property $\langle a \rangle_\mu F$, but it also increases the number of process copies that must be made.

An important issue for future studies is that of the *cost* of a test—e.g., measured by the number of basic experiments (pressure of buttons) required when executing the test. Obviously, when testing for a property with a given level of significance, we would prefer to use a test with lowest possible cost. Dually—a situation which might be more realistic—given an upper bound of the cost, what test will ensure the highest level of significance. Closely related to the cost of tests is their *informativeness*. Loosely, we may consider a test t as being more informative than a test t' if any conclusion that can be derived from observations of t' also may be derived from observations of t. Obviously, we would expect more informative tests also to be more costly. An interesting problem for the future would be to axiomatize the information ordering on tests.

Also, it would be interesting to investigate the testability of other specification formalisms than the modal logics we have considered (e.g., temporal logic and process algebra). This problem is a main topic of a forthcoming thesis (Skou, 1990) by one of the authors.

Recent work by Bloom and Meyer (1989) further shows that if processes p and q are bisimilar, then there is an assignment of probabilities to the transitions of p and q yielding probabilistically bisimilar processes p' and q'.

8. Appendix

In this appendix we give the proofs for the theorems presented in the main paper.

Theorem 8.1 (Theorem 4.5). *Let $\mathscr{P} = (\mathrm{Pr}, \mathrm{Act}, \mathrm{Can}, \mu)$ be a probabilistic transition system satisfying the minimal probability assumption. Then two processes are $\frac{2}{3}$-bisimilar just in case they satisfy exactly the same LML formulas.*

The proof is done by showing that $p \models F \Rightarrow q \models F$ for any formula F if and only if (p, q) is contained in some $\frac{2}{3}$-bisimulation.

$\Leftarrow$ (If):

This follows from induction on the structure of formulas in LML. We only consider the case $\langle a \rangle F$.

So assume that (p, q) is contained in some $\frac{2}{3}$-bisimulation R and assume also that $p \models \langle a \rangle F$. Then there is a p' such that $p \xrightarrow{a} p'$ and $p' \models F$. By the assumption $((p, q) \in R)$ follows the existence of a q' such that $q \xrightarrow{a} q'$ and $(p', q') \in R$, and from (Induction Hypothesis) follows $q' \models F$, which in turn implies $q \models \langle a \rangle F$.

$\Rightarrow$ (Only if):

We show that the relation $R = \{(p, q) \mid \forall F \in \text{LML}.\, p \models F \Rightarrow q \models F\}$ is closed under the definition for $\frac{2}{3}$-bisimulation.

So assume $p \xrightarrow{a} p'$ and consider the finitely many a-derivatives of q, i.e., the processes $(q_1, ..., q_k)$. We want to show that $(p', q_j) \in R$ for some j and we assume that this does not hold for any j. Then there are formulas $F_1, ..., F_k$, where $p \models F_j$ but $q_j \not\models F_j$, and therefore $p \models \langle a \rangle (F_1 \wedge \cdots \wedge F_k)$, whereas $q \not\models \langle a \rangle (F_1 \wedge \cdots \wedge F_k)$, which contradicts $(p, q) \in R$.

Now assume $q \xrightarrow{a}$. Then $q \not\models [a]\text{ff}$ and therefore $p \not\models [a]\,\text{ff}$ which means $p \xrightarrow{a}$.

THEOREM 8.2 (Theorem 5.1). *The formulas of Hennessy–Milner Logic are testable.*

The proof is by induction on the structure of formulas.

ff:

Choose $t = \omega$, $E_\omega = \varnothing$. Then $p \models \text{ff}$ is always false, and $p \not\models \text{ff} \Rightarrow P_{\omega, p}(E_\omega) = 0$.

$\langle a \rangle F$:

Choose $t = (a.t_F)^N$, $E_{\langle a \rangle F} = \{(e_1, ..., e_N) \mid \exists i, e_i = 1_a : e_i', e_i' \in E_F\}$. Suppose now $p \models \langle a \rangle F$. Because of the image-finiteness there is an $\varepsilon > 0$ such that $\mu_{p,a}(p') \geqslant \varepsilon$ for all p', $p \xrightarrow{a} p'$. For a given e_i resulting from $a.t_F$, the probability that it has form $1_a : e_i', e_i' \in E_F$ is then at least $\varepsilon(1 - \delta_F)$, where δ_F is chosen arbitrarily (implying by IH a t_F such that $P_{t, p'}(E_F) \geqslant (1 - \delta_F)$). So we have $P_{t, p}(E_{\langle a \rangle F}) \geqslant 1 - (1 - \varepsilon(1 - \delta_F))^N$. In the same way it can be argued that $p \not\models \langle a \rangle F$ implies that $P_{t, p}(E_{\langle a \rangle F}) \leqslant 1 - (1 - \delta_F)^N$.

In the case of $p \models \langle a \rangle F$ we assume that all choices of δ_F are made so that $\delta_F \leqslant \frac{1}{2}$ and we can then restate the two inequalities as

1. $p \models \langle a \rangle F \Rightarrow P_{t,p}(E_{\langle a \rangle F}) \geqslant 1-(1-\varepsilon/2)^N$
2. $p \not\models \langle a \rangle F \Rightarrow P_{t,p}(E_{\langle a \rangle F}) \leqslant 1-(1-\delta_F)^N$.

For a given δ we can first choose N such that $(1-(1-\varepsilon/2))^N \geqslant 1-\delta$. Then we can chose δ_F, t_F by Induction Hypothesis such that $1-(1-\delta_F)^N \leqslant \delta$.

$[a]F$:

Choose $t=(a.t_F)^N$, where t_F is chosen according to some arbitrary δ_F, and $E_{[a]F}=\{(e_1, ..., e_N) \mid \forall i: e_i = 1_a : e_i', e_i' \in E_F\}$, and suppose $p \models [a]F$. For all p', $p \xrightarrow{a} p'$ we have $P_{t_F, p'}(E_F) \geqslant 1-\delta_F$. Therefore $P_{t,p}(E_{[a]F}) \geqslant (1-\delta_F)^N$.

Suppose also $p \not\models [a]F$. This means that there is at least one p', $p \xrightarrow{a} p'$ such that $p' \not\models F$. Using the arguments of the previous case we then have $P_{t,p}(E_{[a]F}) \leqslant (1-\varepsilon(1-\delta_F))^N$. As before we assume $\delta_F \leqslant \varepsilon/2$, and we can then restate the inequalities as follows:

1. $p \models [a]F \Rightarrow P_{t,p}(E_{[a]F}) \geqslant (1-\delta_F)^N$
2. $p \not\models [a]F \Rightarrow P_{t,p}(E_{[a]F}) \leqslant (1-\varepsilon/2)^N$.

As before we can then for given δ choose an N such that $(1-\varepsilon/2)^N \leqslant \delta$ and thereby t_F such that $(1-\delta_F)^N \geqslant 1-\delta$.

$F_1 \wedge F_2$:

Use $t=(t_{F_1}, t_{F_2})$ and $E_F=\{(e_1, e_2) \mid e_1 \in E_{F_1} \wedge e_2 \in E_{F_2}\}$. We then have

$$p \models F_1 \wedge F_2 \Rightarrow P_{t,p}(E_F) \geqslant (1-\delta_{F_1}-\delta_{F_2})$$
$$p \not\models F_1 \wedge F_2 \Rightarrow P_{t,p}(E_F) \leqslant \delta_{F_1}\delta_{F_2}.$$

For given δ we can therefore choose a test satisfying the general conditions for testability.

$F_1 \vee F_2$:

Use $t=(t_{F_1}, t_{F_2})$ and $E_F=\{(e_1, e_2) \mid e_1 \in E_{F_1} \vee e_2 \in E_{F_2}\}$. The arguments then follow along the lines of the previous case.

In the following proof we shall use the well-known Chebyshev's inequality, which estimates the probability of the deviation between a random variable and its mean (see (Chung, 1974)):

LEMMA 8.3. *Let X be a random variable having mean ρ and variance σ^2. Then the following holds for any $k>0$:*

$$P(|X-\rho| \geqslant k \cdot \sigma) \leqslant \frac{1}{k^2}.$$

THEOREM 8.4 (Theorem 6.2). *The formulas of Probabilistic Modal Logic are testable if the minimal deviation assumption is satisfied.*

The proof is by induction on the structure of formulas. We only show the case $\langle a \rangle_\mu F$ since the remaining cases are shown in the testability proof of HML (Theorem 5.1).

Consider now a test of form $a.t_F$, where t_F (inductively) is assumed to be a test for F with set of evidence E_F and level of significance δ_F. We want to find the probability of obtaining an observation within the set $E = 1_a : E_F$. So assume that the investigated process p enjoys the property $\langle a \rangle_\mu F$. Then there is a $\mu_0 \geqslant \mu$ such that $p \xrightarrow{a}_{\mu_0} S$, where $p' \in S$ implies $p' \models F$. Because of the minimal deviation assumption, μ_0 has form $M \cdot \varepsilon$, where $M \geqslant \lceil \mu/\varepsilon \rceil$. Therefore we have

$$
\begin{aligned}
P_{a.t_F, p}(E) &= P_{a.t_F, p}(1_a : E_F) \\
&= \sum_{p'} \mu_{p,a}(p') \cdot P_{t_F, p'}(E_F) \\
&\geqslant \sum_{p' \in S} \mu_{p,a}(p') \cdot P_{t_F, p'}(E_F) \\
&\geqslant \lceil \mu/\varepsilon \rceil \cdot \varepsilon \cdot (1 - \delta_F) \\
&\triangleq \rho_0.
\end{aligned}
$$

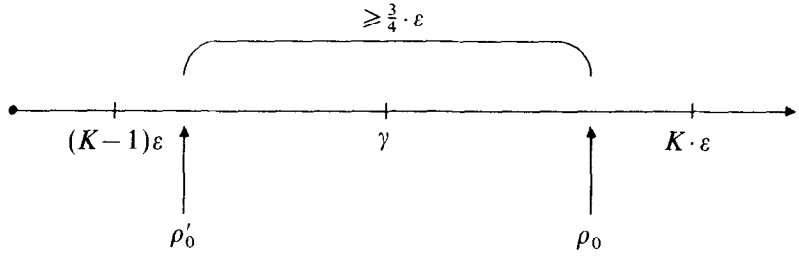
Now assume that p does not enjoy $\langle a \rangle_\mu F$. We can then divide the a-derivations of p into sets S_1, S_2, where S_1 consists of those p' for which $p' \models F$ holds. So we have $p \xrightarrow{a}_{\mu_0} S_1$, where $\mu_0 < \mu$. As before, μ_0 has form $N \cdot \varepsilon$, where now $N \leqslant \lceil \mu/\varepsilon \rceil - 1$. Hence we can estimate the probability of E:

$$
\begin{aligned}
P_{a.t_F, p}(E) &= P_{a.t_F, p}(1_a : E_F) \\
&= \sum_{p'} \mu_{p,a}(p') \cdot P_{t_F, p'}(E_F) \\
&= \sum_{p' \in S_1} \mu_{p,a}(p') \cdot P_{t_F, p'}(E_F) + \sum_{p' \in S_2} \mu_{p,a}(p') \cdot P_{t_F, p'}(E_F) \\
&\leqslant \lceil \mu/\varepsilon \rceil - 1 \cdot \varepsilon + \delta_F \\
&\triangleq \rho_0'.
\end{aligned}
$$

Let K denote $\lceil \mu/\varepsilon \rceil$. We can choose δ_F (and thereby t_F and E_F) such that

$\rho_0 - \rho_0'$ $(= \varepsilon - \delta_F \cdot (1 + K \cdot \varepsilon))$ comes arbitrary close to ε, say $\frac{3}{4}\varepsilon$. Letting γ denote $(K - \frac{1}{2}) \cdot \varepsilon$ we have (see illustration below)

$$\rho_0' + \tfrac{1}{4} \cdot \varepsilon < \gamma < \rho_0 - \tfrac{1}{4} \cdot \varepsilon \tag{1}$$



Consider then a test of form $(a.t_F)^N$, where t_F is the test defined previously. We now define the set of evidence

$$E_{\langle a \rangle_\mu F} = \left\{ (1_a : e_1, \ldots, 1_a : e_N) \;\middle|\; \frac{\#e_i \in E_F}{N} \geqslant \gamma \right\};$$

i.e., we require that a certain minimum proportion of the component observations e_i in the observation $(1_a : e_1, \ldots, 1_a : e_N)$ belong to E_F. For any probabilistic process p, $(\#e_i \in E_F)$ is a random variable having a binomial distribution (see (Chung, 1974)) $\mathscr{B}(N, P_{a.t_F,p}(1_a : E_F))$; i.e., it has mean $N \cdot P_{a.t_F,p}(1_a : E_F)$ and variance $N \cdot P_{a.t_F,p}(1_a : E_F) \cdot (1 - P_{a.t_F,p}(1_a : E_F))$. Thus the random variable $(\#e_i \in E_F)/N$ has mean $\rho_p = P_{a.t_F,p}(1_a : E_F)$ and variance $\rho_p \cdot (1 - \rho_p)/N$.

We denote this random variable X_p^N, and from the estimates above it follows that $\rho_p \leqslant \rho_0'$ whenever $p \not\models \langle a \rangle_\mu F$ and $\rho_0 \leqslant \rho_p$ whenever $p \models \langle a \rangle_\mu F$.

We can now estimate the probabilities of evidence for $t = (a.t_F)^N$, i.e., $P_{t,p}(E_{\langle a \rangle_\mu F})$.

In case $p \models \langle a \rangle_\mu F$ we have

$$\begin{aligned}
P_{t,p}(E_{\langle a \rangle_\mu F}) &= P_{t,p}(X_p^N \geqslant \gamma) \\
&= 1 - P_{t,p}(X_p^N < \gamma) \\
&\geqslant 1 - P_{t,p}\left(|X_p^N - \rho_p| > \frac{1}{4} \cdot \varepsilon \right) \quad \left(\text{because } \rho_p - \gamma \geqslant \rho_0 - \gamma \geqslant \frac{1}{4} \cdot \varepsilon \right) \\
&\geqslant 1 - \left(\frac{4}{\varepsilon} \right)^2 \cdot \frac{\rho_p (1 - \rho_p)}{N} \qquad \text{(because of Chebyshev's inequality)} \\
&\geqslant 1 - \left(\frac{4}{\varepsilon} \right)^2 \cdot \frac{1}{N}.
\end{aligned}$$

In case $p \not\models \langle a \rangle_\mu F$ we have

$$\begin{aligned} P_{t,p}(E_{\langle a \rangle_\mu F}) &= P_{t,p}(X_p^N \geqslant \gamma) \\ &\leqslant P_{t,p}\left(|X_p^N - \rho_p| > \frac{1}{4} \cdot \varepsilon\right) \qquad \left(\text{because } \gamma - \rho_p \geqslant \rho_0' - \rho_p \geqslant \frac{1}{4} \cdot \varepsilon\right) \\ &\leqslant \left(\frac{4}{\varepsilon}\right)^2 \cdot \frac{\rho_p(1-\rho_p)}{N} \qquad \text{(Chebyshev's inequality)} \\ &\leqslant \left(\frac{4}{\varepsilon}\right)^2 \cdot \frac{1}{N}. \end{aligned}$$

For a given δ we first choose δ_F such that (1) holds and thereby N such that $\delta \geqslant (4/\varepsilon)^2 \cdot 1/N$. Then we have

$$p \models \langle a \rangle_\mu F \Rightarrow P_{t,p}(E_{\langle a \rangle_\mu F}) \geqslant 1 - \left(\frac{4}{\varepsilon}\right)^2 \cdot \frac{1}{N} \geqslant 1 - \delta$$

$$p \not\models \langle a \rangle_\mu F \Rightarrow P_{t,p}(E_{\langle a \rangle_\mu F}) \leqslant \left(\frac{4}{\varepsilon}\right)^2 \cdot \frac{1}{N} \leqslant \delta.$$

In order to prove Theorem 6.4 we define the following notion of a dual formula F^{D} for a given formula F in PML:

DEFINITION 8.5. The dual formula F^{D} for a given formula F in PML is defined structurally as follows:

$$\begin{aligned} \mathrm{tt}^{\mathrm{D}} &= \mathrm{ff} \\ \mathrm{ff}^{\mathrm{D}} &= \mathrm{tt} \\ (F \wedge G)^{\mathrm{D}} &= F^{\mathrm{D}} \vee G^{\mathrm{D}} \\ (F \vee G)^{\mathrm{D}} &= F^{\mathrm{D}} \wedge G^{\mathrm{D}} \\ \Delta_a^{\mathrm{D}} &= \langle a \rangle_1 \mathrm{tt} \\ (\langle a \rangle_\mu F)^{\mathrm{D}} &= \Delta_a \vee \langle a \rangle_{1-(\lceil \mu/\varepsilon \rceil - 1)\cdot\varepsilon} F^{\mathrm{D}}. \end{aligned}$$

The property $(\langle a \rangle_\mu F)^{\mathrm{D}}$ is motivated as follows: If Δ_a holds, obviously $\langle a \rangle_\mu F$ does not hold, and if $\langle a \rangle_x F^{\mathrm{D}}$ holds for some x, this clearly excludes the fulfillment of $\langle a \rangle_\mu F$, when x is strictly larger than $1 - \mu$. It is easily seen that $1 - (\lceil \mu/\varepsilon \rceil - 1) \cdot \varepsilon$ is such an x.

A given formula and its dual are related by the following duality lemma, which we state without its simple proof:

LEMMA 8.6.

$$p \models F \Rightarrow p \not\models F^{\mathrm{D}}$$
$$p \not\models F \Rightarrow p \models F^{\mathrm{D}}$$

THEOREM 8.7 (Theorem 6.4). *Let* $\mathscr{P} = (\mathrm{Pr}, \mathrm{Act}, \mathrm{Can}, \mu)$ *be a probabilistic transition system satisfying the minimal deviation assumption. Then two processes are probabilistic bisimilar just in case they satisfy exactly the same PML formulas.*

$\Rightarrow$(Bisimilar processes satisfy the same formulas):

The proof is by induction on the structure of formulas. We show only the case $\langle a \rangle_\mu F$ since the remaining cases are trivial.

So assume $p \equiv_P q$ and $p \models \langle a \rangle_\mu F$. Then there is a set $S \subseteq \mathrm{Pr}$ with $p \xrightarrow{a}_{\mu'} S$, $\mu' \geqslant \mu$ and $p'' \models F$ for all $p'' \in S$.

Consider now $S' = \bigcup \{T \in \mathrm{Pr}/\equiv_P \mid T \cap S \neq \varnothing\}$. Then for every $p' \in S'$ there is a process $p'' \in S$ with $p' \equiv_P p''$, and since all processes in S satisfy the formula F, we can from the Induction Hypothesis conclude that $p' \models F$ for all $p' \in S'$.

Since $S \subseteq S'$, obviously $p \xrightarrow{a}_{\mu''} S'$ with $\mu'' \geqslant \mu'$.

Also, since S' is a union of equivalence classes under $\equiv_P$, it follows immediately from $p \equiv_P q$ that $q \xrightarrow{a}_{\mu''} S'$, showing that $q \models \langle a \rangle_\mu F$. By symmetry of $\equiv_P$ it then follows that p and q satisfy the same formulas.

$\Leftarrow$ (Processes satisfying the same formulas are bisimilar):

Consider the relation $R = \{(p, q) \mid p$ and q satisfy the same PML formulas$\}$. We want to show that R is closed under the definition of $\equiv_P$. (Obviously R is an equivalence.)

So let pRq and asume $p \xrightarrow{a}_{\mu_0} S_0$, $S_0 \in \mathrm{Pr}/R$. Furthermore let $q \xrightarrow{a}_{\mu_1} q_1, \ldots, q \xrightarrow{a}_{\mu_j} q_j$, $q \xrightarrow{a}_{\mu_{j+1}} q_{j+1}, \ldots, q \xrightarrow{a}_{\mu_k} q_k$ be the a-derivatives of q (of which there are finitely many due to the minimal deviation property). The derivatives have been ordered such that $q_1, \ldots, q_j \in S_0$ and $q_{j+1}, \ldots, q_k \notin S_0$.

Then, because of the duality lemma, we can find PML formulas $F_{j+1}, \ldots, F_k$ such that $S_0 \models F_i$ but $q_i \not\models F_i$ for all $i = j+1, \ldots, k$ (extending $\models$ to classes under R).

Now assume $\sum_{i=1}^{j} \mu_i < \mu_0$. Then $p \models \langle a \rangle_{\mu_0}(\bigwedge_{i=j+1}^{k} F_i)$ but $q \not\models \langle a \rangle_{\mu_0}(\bigwedge_{i=j+1}^{k} F_i)$ contradicting the assumption pRq. Thus $\mu_0 \leqslant \sum_{i=1}^{j} \mu_i$ and therefore $q \xrightarrow{a}_{\mu_0'} S_0$ with $\mu_0' \geqslant \sum_{i=1}^{j} \mu_i \geqslant \mu_0$. By symmetry of R, it follows that $\mu_0 \geqslant \mu_0'$ and hence $\mu_0 = \mu_0'$.

This proves that R is a probabilistic bisimulation.

THEOREM 8.8 (Proof of Theorem 6.5). *Let* $\mathscr{P} = (\mathrm{Pr}, \mathrm{Act}, \mathrm{Can}, \mu)$ *be a*

probabilistic transition system satisfying the minimal deviation assumption. Then $p \equiv_P q$ *just in case* $P_{t,p}(e) = P_{t,q}(e)$ *for all tests* t *and observations* $e \in \mathcal{O}_t$.

$\Rightarrow$ ($p \equiv_P q$ implies identical distributions):

The proof is by induction on the structure of tests and we consider only the case $a.t$, since the other cases are trivial.

If $p \not\xrightarrow{a}$ we have by the assumption that $q \not\xrightarrow{a}$, which implies $P_{a.t,p}(0_a) = P_{a.t,q}(0_a) = 1$ and $P_{a.t,p}(1_a : e) = P_{a.t,q}(1_a : e) = 0$ for any observation $e \in O_t$.

If $p \xrightarrow{a}$, also $q \xrightarrow{a}$ holds and therefore $P_{a.t,p}(0_a) = P_{a.t,q}(0_a) = 0$. Now let $S_1, \ldots, S_k$ denote the (finitely many) equivalence classes under $\equiv_P$ which intersect $\{p' \mid p \xrightarrow{a} p'\}$. From the Induction Hypothesis it follows that for any observation $e \in O_t$ the processes in each S_j yield identical probability, i.e., $P_{t,p'}(e) = k_j$ for all $p' \in S_j$. Since $p \xrightarrow{a}_{\mu_j} S_j \Leftrightarrow q \xrightarrow{a}_{\mu_j} S_j$ holds for each S_j, we have for any $e \in O_t$

$$\begin{aligned} P_{a.t,p}(1_a : e) &= \sum_{p'} \mu_{p,a}(p') \cdot P_{t,p'}(e) \\ &= \sum_{S_j} \mu_j \cdot k_j \\ &= \sum_{q'} \mu_{q,a}(q') \cdot P_{t,q'}(e) \\ &= P_{a.t,q}(1_a : e). \end{aligned}$$

$\Leftarrow$ (Identical distributions imply $p \equiv_P q$):

The proof is by contradiction; i.e., we assume that $p \not\equiv_P q$. Theorem 6.4 ensures that we then can find a property F of PML which is enjoyed by p, but not by q. As PML is testable, we can find a test t_F and a set $E_F \subseteq O_{t_F}$ such that $P_{t_F,p}(E_F) \geq \frac{3}{4}$ and $P_{t_F,q}(E_F) \leq \frac{1}{4}$ i.e. we have $P_{t_F,p}(E_F) \neq P_{t_F,q}(E_F)$, contradicting the assumption that p and q induce identical probability distributions for any test.

ACKNOWLEDGMENTS

This work—owing inspiration to the framework of Samson Abramsky (1987)—has been carried out as part of the TAU-project, a project supported by the FTU-program under the Danish Research Council. We thank Susanne Christensen, Steffen Lauritzen, Søren Lundbye Christensen, and Aage Nielsen from the Statistics Group of our department for the many helpful discussions on the fundamentals of statistics and hypothesis testing. Also thanks to Krik for making the pictures in this paper.

RECEIVED June 28, 1989

REFERENCES

ABRAMSKY, S. (1987), Observation equivalence as a testing equivalence, *Theoret. Comput. Sci.* **53**, 225.

BLOOM, B., AND MEYER, A. (1989), A remark on bisimulation between probabilistic processes, *in* "Logic at Botik '89" (A. Meyer and M. Taitslin, Eds.), pp. 26–40, Lecture Notes in Computer Science, Vol. 363, Springer-Verlag, Berlin/New York.

BLOOM, B., ISTRAIL, S., AND MEYER, A. (1988), Bisimulation can't be traced: Preliminary report, *in* "Proceedings 15th ACM POPL," pp. 229–239.

CHUNG, K. L. (1974), "Elementary Probability Theory with Stochastic Processes," Springer-Verlag, Berlin/New York.

COX, D., AND HINKLEY, D. (1974), "Theoretical Statistics," Chapman & Hall, London.

DE NICOLA, R. AND HENNESSY, M. (1983), Testing equivalences for proceses, *Theoret. Comput. Sci.* **34**, 83.

DIJKSTRA, E. (1972), Notes on structured programming, *in* "Structured Programming" (O. Dahl, E. Dijkstra, and C. Hoare, Eds.), pp. 1–82, Academic Press, New York.

HENNESSY, M. AND MILNER, R. (1985), Algebraic laws for nondeterminism and concurrency, *J. Assoc. Comput. Mach.*, 137.

HUGHES, G., AND CRESSWELL, M. (1972), "An Introduction to Modal Logic," Methuen, London.

LARSEN, K. G. (1986), "Context-Dependent Bisimulation Between Processes," Ph.D. thesis, University of Edinburgh.

LARSEN, K. G. (1988), Proof systems for Hennessy–Milner logic with recursion, *in* "CAAP '88" (M. Dauchet and M. Nivat, Eds.), pp. 215–230, Lecture Notes in Computer Science, Vol. 299, Springer-Verlag, Berlin/New York.

MILNER, R. (1980), "Calculus of Communicating Systems," Lecture Notes in Computer Science, Vol. 92, Springer-Verlag, Berlin/New York.

MILNER, R. (1981), Modal characterization of observable machine behaviour, *in* "CAAP '81 Proceedings" (G. Astesiano and C. Böhm, Eds.), Lecture Notes in Computer Science, Vol. 112, Springer-Verlag, Berlin/New York.

MILNER, R. (1963), Calculi for synchrony and asynchrony, *Theoret. Comput. Sci.* **25**, 267.

PARK, D. (1981), Concurrency and automata on infinite sequences, *in* "5th GI Conference" (P. Deussen, Ed.), pp. 167–183, Lecture Notes in Computer Science, Vol. 104, Springer-Verlag, Berlin/New York.

PHILLIPS, I. (1986), Refusal testing, *in* "ICALP '86" (L. Kott, Ed.), pp. 304–313, Lecture Notes in Computer Science, Vol. 226, Springer-Verlag, Berlin/New York.

PLOTKIN, G. (1981), A structural approach to operational semantics, FN 19, DAIMI, Aarhus, University.

PNUELI, A. (1985), Linear and branching structures in the semantics and logics of reactive systems, *in* "ICALP '85" (W. Brauer, Ed.), pp. 15–32, Lecture Notes in Computer Science, Vol. 194, Springer-Verlag, Berlin/New York.

SKOU, A. (1990). "Validation of Concurrent Processes, with Emphasis of Testing," Ph.D. thesis, Aalborg University Centre.